QH
301
B615

Vol. 44, Parts 1 and 2 June 1957

UNIVERSITY OF MICHIGAN
JUN 2 1957
SCIENCE LIBRARY

QR

BIOMETRIKA

FOUNDED BY

W. F. R. WELDON, FRANCIS GALTON AND KARL PEARSON

MANAGING EDITOR

E. S. PEARSON

ASSOCIATE EDITOR

M. G. KENDALL

ISSUED BY

THE BIOMETRIKA OFFICE, UNIVERSITY COLLEGE, LONDON

PRINTED AT THE UNIVERSITY PRESS, CAMBRIDGE

[*Issued* 4 *June* 1957]

This volume of *Biometrika* is published with the co-operation of

F. N. DAVID — D. G. KENDALL
J. DURBIN — D. V. LINDLEY
M. J. R. HEALY — R. L. PLACKETT
J. B. S. HALDANE

A volume containing about 500 pages will be published annually in two half-yearly issues, appearing in June and December.

Papers for publication should be sent either to

PROFESSOR E. S. PEARSON

Department of Statistics, University College, London, W.C. 1

or if more convenient to

PROFESSOR M. G. KENDALL

London School of Economics, Houghton Street, London, W.C. 2

It is a condition of publication in *Biometrika* that the paper shall not already have been issued elsewhere, and will not be reprinted without leave of the Editors.

Contributors receive 25 copies of their papers free. Joint authors 15 copies each. Order forms for separates are sent to authors with proofs of their papers.

The Subscription price, payable in advance, is:

Inland: 45*s.* net per volume Abroad: 54*s.* net per volume
including packing and postage

BACK ISSUES

Volumes 20–43

These may be obtained from the BIOMETRIKA OFFICE, at the following prices:

Vols. 20–38: £5 per volume Vols. 39–43: £3. 3*s.* per volume
packing and postage 2s. per volume

Bound volumes: £1 extra per volume Binding cases: 10*s.* each

Cheques should be made payable to *Biometrika*, crossed "a/c BIOMETRIKA TRUST" and sent to

THE SECRETARY
BIOMETRIKA OFFICE
UNIVERSITY COLLEGE
GOWER STREET, LONDON, W.C. 1

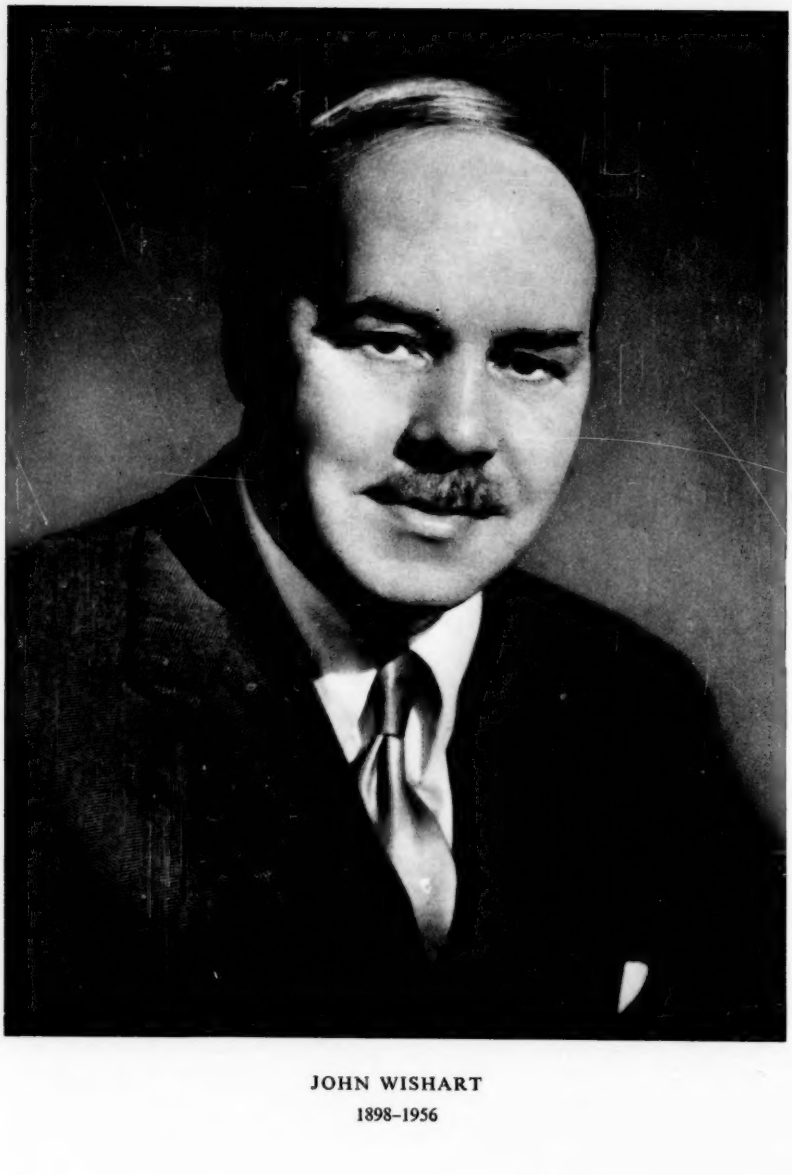
to whom all orders for series, single copies and offprints should be addressed.

All foreign cheques must be drawn in sterling and on a Bank having a London Agency.

Volumes 1–19

Permission for reprinting has been granted to Messrs Wm. Dawson & Sons Ltd. Volumes 1–13 are now ready for distribution and Volumes 14–19 are in course of preparation. Would librarians and others wishing to have copies please place their orders now with:

WM. DAWSON & SONS LTD.,
4 DUKE STREET, MANCHESTER SQUARE,
LONDON, W. 1



JOHN WISHART
1898–1956

VOLUME 44, PARTS 1 AND 2 JUNE 1957

JOHN WISHART

1898–1956

There are probably few statisticians who have had more friends scattered across the world than had John Wishart. Many of these friendships were made in the course of his work as a teacher of statistical method to practical agriculturalists overseas. It was in the middle of teaching work of this kind, which perhaps he had come to enjoy more than any other and in which he undoubtedly excelled, that he died suddenly on 14 July 1956 while bathing at Acapulco on the Pacific coast of Mexico. A pioneering visit to Nanking University in 1934 had been followed after the war by visits to Spain in 1947, to the United States in 1949, to India in 1954 and then, in this last year, to Mexico, where he was taking a leading part in the work of the Training Centre in Experimental Design arranged by the United Nations Food and Agricultural Organization.

John was the second of the four sons of John and Elizabeth Wishart; he was born in Montrose on 28 November 1898, the family moving to Perth in 1900. After schooling at Perth Academy he went to Edinburgh University in 1916, finally obtaining a 1st Class Honours Degree in Mathematics and Natural Philosophy in 1922. His university career was broken by two years' service in the Army (1917–19), when he was a subaltern in the 7th Battalion of the Black Watch, and saw active service in France during the last few months of the war. His four years at the university included a teacher's training course at Moray House, which led him on to his first post as a mathematics master at the West Leeds High School (1922–4). Here he met Olive Birdsall, whom he married in 1924. There were two sons of the marriage.

When on the look-out for mathematical assistants, Karl Pearson had on several occasions asked E. T. Whittaker for a suitable name among his past Edinburgh students, and it was with a recommendation of this kind that Wishart came to University College, London, in the autumn of 1924 to start on a statistical career. The post which he filled was that of Research Assistant to Pearson, the funds being supplied by the Department of Scientific and Industrial Research.

The computation of the *Tables of the Incomplete Gamma Function* had been finished, the Tables being issued in 1922, and Pearson was already preparing the ground for his next great computational undertaking, that of the *Tables of the Incomplete Beta Function*. It followed that one of Wishart's main tasks on arriving in London was to get this work on the Beta function under way. I cannot recall how far he got with the tabling programme, which was not finally completed till 1932, but there is no doubt that his first years of apprenticeship in statistics involved him in much hard computing labour. Three of his earliest papers, too, published in *Biometrika* (**1, 2, 5***) were concerned with methods of approximation to the incomplete Beta function. Both the Gamma and Beta distributions and their practical uses were a continuing source of interest to him, and one of his last papers published in 1956 (**77**) was concerned with a new method of approximating to the integral of the former.

* Here, and below, the numbers are those given in order of date in the Bibliography on pp. 6–8 below.

The portrait facing this page was taken by Bassano in 1951.

I do not think that Wishart regretted his hard training in computation which had started, without machine aid, in Whittaker's mathematical laboratory. Later on he was to play an active part on the British Association's Mathematical Tables Committee (1928–48), of which he was for some years the secretary, and *Biometrika* owes him much for the help, advice and criticism which he gave in the recasting of the *Biometrika Tables for Statisticians*.

During his three years at University College Wishart attended the two main courses of lectures on mathematical statistics which Pearson was giving annually, and he learned how to handle mathematical problems and put together a scientific paper, but the terms of his appointment as research assistant precluded any form of teaching. He was a teacher by temperament and training, and it was, perhaps, the prospect of starting some statistical classes at a new centre, rather than any marked advance in salary, which led him in the autumn of 1927 to accept a post as Mathematical Demonstrator at the Imperial College of Science.

He had scarcely settled in his new post, however, when a second great opportunity came his way and he found himself at the beginning of 1928 appointed as Statistical Assistant under R. A. Fisher at Rothamsted. Fisher's scientific output had at this juncture reached a remarkable tempo, and Wishart had the good fortune to see at close quarters and to share in the development of much new theory and in the trying out of that theory in practical experimentation.

His very considerable number of publications during the years 1928–32 illustrate well Wishart's fields of research. On the more mathematical side, with Fisher's encouragement, he broke new ground in 1928 with the derivation of the generalized product-moment distribution (**8**) which was to play an important part in the development of multivariate analysis. A number of his other papers followed from this beginning (e.g. (**10, 13**)). Other papers (**9, 11, 23, 29** were concerned with the properties of the distribution of the multiple correlation coefficient and allied distributions which Fisher had derived between 1924 and 1928, using geometrical methods. Wishart took a hand, too, in Fisher's work of showing how the impasse in the theory of sampling moments which seemed to have been reached in the 1920's, could be broken through by a systematic use of k-statistics and the methods of combinatorial analysis (**12, 21, 24**). I well remember how my first gleam of understanding on partitional problems came from Wishart's demonstration with match-stick patterns in his sitting-room at Harpenden.

At the same time Wishart was taking his full share in the Statistical Department's work of servicing the experimental programme at Rothamsted. Joint papers such as those with Clapham (**15**), Fisher (**17**) and Allan (**18**) were pioneer efforts, written to illustrate how the new techniques were working out in practice. In addition, he wrote or shared in writing a number of simple expository papers (e.g. (**14, 20, 22, 26**)). His power of putting things clearly to his non-mathematical colleagues was certainly appreciated. As Sir John Russell wrote shortly after Wishart's death:

> It seems only the other day that he first came to Rothamsted to help in the development of what was then a new and untried subject, of the value of which many were yet uncertain. He threw himself into the work with energy and enthusiasm, mastered its intricacies and having the gift of exposition was able to explain its difficulties to the various young people at Rothamsted who wanted to learn something about it but feared it was beyond their grasp. But he succeeded and he helped many....We all regretted his departure to Cambridge, but of course we knew it was a bigger field giving him more scope.

In 1931 Udny Yule decided to give up his post of University Lecturer in Statistics at Cambridge which he had held since 1912, and a committee of the University recommended some reorganization of the teaching work. A Reader* was to be appointed in the Faculty of Agriculture, who should also do some teaching in the Faculty of Mathematics, while Economic Statistics was to be the concern of a separate Lecturer. Wishart was appointed to take up the Readership in October 1931; there was no one more suited for the post. He had behind him 3½ years experience of applying the new statistical methods in agricultural experimentation and of explaining their meaning to the non-mathematician; he had received an apprenticeship in mathematical statistics under both Karl Pearson and R. A. Fisher; he had a shrewd, objective judgement which enabled him to keep clear of controversy and to combine what he thought good from the older and newer schools of thought; and, finally, he was ready to take any amount of trouble in developing lecture and practical courses for new types of student audience.

He was given accommodation in the School of Agriculture, where in addition to his own room he had a larger 'laboratory' containing some half dozen desks for post-graduate students. He began his teaching with a general course on statistical methods offered in the Faculty of Agriculture and a course on mathematical statistics which could be offered for Schedule B of the Mathematical Tripos. There was also an optional practical class associated with the second course. In the eight years which intervened before the war, the number of mathematical students who attended his Tripos course and in many cases stayed on for post-graduate work increased steadily. This drawing of Cambridge-trained mathematicians into the field of statistics had a profound effect on the rate of development of the subject in this country, and it was Wishart who played the essential part in achieving this result.

Impressions of these days left on his early students are interesting. As is inevitably the case when statistics teaching can only be fitted in as one subject in a mathematical syllabus, the Schedule B course was rather overcrowded. It was in the more leisurely conditions of full-time post-graduate study that Wishart's teaching was most effective. A characteristic aspect of the research laboratory has been described by W. G. Cochran:

> In those days he believed in his students keeping office hours. When he assigned me a desk in the Lab., he told me that he expected me to be sitting at the desk most of the day when not in class. He instructed me to do three hours computing a day on a table of the 1 % level of z to 7 decimal places, while Fisher in the meantime was computing the 5 % table. Having anticipated a free and easy life as a graduate student, punctuated of course by periods of esoteric thinking when the spirit moved me, I didn't much like either the office hours or the computing, but I don't think they did me any harm.

In asking post-graduates to keep regular hours for statistical work and later on in expecting somewhat similar regularity from his lecturers, he was following, perhaps unconsciously, Pearson's tradition established at University College. It was a routine less readily accepted at Cambridge, especially on the mathematical side. Another custom in which he followed K. P. was in making daily rounds of the laboratory to see how his students were getting on and to discuss any difficulties with them.

H. O. Hartley was a student a year or two after Cochran, and writes:

> John Wishart's strength as a teacher of applied statistics was two-sided. He had an amazing gift to inspire a very theoretical mathematician, like myself, to learn something about applications and to bring him into contact with the agriculturalist and biologist. He was also successful, although less so, in converting the latter to see the need of mathematical statistics. He would walk into the laboratory

* Yule was made Reader for the last few months of his appointment.

and look over your shoulder as you were struggling with an analysis of variance, point to a figure and say 'this one looks wrong', and would usually be infallible in this. He had a computational flair, a great gift for approximations and took a great interest in computing methods, tables and machines. In short he was an ideal teacher of statistical analysis, but not quite as convincing as a teacher of statistical inference. He had a good humoured gift to learn from failure and tell the stories against himself for others to learn from.

A feature of this Cambridge development in the 1930's was the location of the statistical laboratory in the School of Agriculture, with agricultural and mathematical students working more or less side by side. 'This was, if you like', M. S. Bartlett points out, 'a reflexion of Wishart's dual qualifications; in those early days at least it was a very happy combination which set the tone for a balanced outlook on the theory of statistics, which was always to be found in his department.'

It has been remarked that Wishart did not produce much original research in these pre-war years, and it is true that if we look at his list of publications, the papers written between 1933 and 1939 were almost entirely of an applied or expository character. But original mathematical research is not the only mark of a good statistician. We should remember that until Bartlett was appointed in 1938 as a mathematical lecturer to help in the statistical work, Wishart was teaching single-handed. At the same time he was acting as consultant in statistical matters in connexion with the experimental work of various Cambridge research institutes. It was a time too when much hard spadework in the way of relatively simple exposition was essential to convince the experimentalists at Cambridge and elsewhere of the value of the new statistical tools. This was a type of work that he found most congenial and he threw himself into it to the full.

The Industrial and Agricultural Research Section of the Royal Statistical Society was founded in 1933 and Wishart was one of the six Fellows of the Society who formed the original organizing committee. He gave the second paper to the section in 1934 on 'Statistics in Agricultural Research' (35), and later in the same year provided a long 'Bibliography of Agricultural Statistics, 1931–33' (36) for volume 1 of the I.A.R.S. Supplement. Indeed, he continued to play an active part in the section's first six years of very vigorous life.

If the particular organization of statistics at Cambridge worked admirably to start with, it did contain the seeds of later trouble. As the agriculturalists became more familiar with statistical method and the routine of experimental design was established, the need for a Reader with Wishart's qualifications within the School of Agriculture became less apparent. The natural development would seem to have been the creation of an independent university statistical unit or department, drawing its recruits from mathematics and serving the university as a whole. But it has never been an easy task to convince a Mathematics Faculty that Mathematical Statistics is a subject with a discipline of its own, worthy of being allowed an independent existence, and Wishart, both before and after the war, found many of the Cambridge mathematicians as conservative in outlook as any in this respect. He found himself between two stools, and though there were no doubt faults on all sides to account for the friction which resulted, his position was not altogether an easy one. As in many another case, when war came he was glad to be able to throw himself into another form of occupation which for a time, at any rate, would free him from more academic battles.

From May 1940 to February 1942 he was in the Army as a Captain employed on Intelligence work, and then, until 1946, he was occupied with statistical work for the Admiralty in London and Bath, working with the rank of Assistant Secretary in the Establishment, the Production and the Naval Personnel Divisions of the Secretary's Department. When

the war ended he hesitated for some time whether to return to Cambridge, but no doubt the great opportunity of developing the work he had started there drew him back to university teaching.

On the mathematical side Wishart was now able to make great progress. Additional members of the Mathematics Faculty were appointed who were to be primarily concerned with statistical teaching, so that by the early 1950's the Reader could call on the help of three lecturers and an assistant lecturer. Part II of the revised Mathematical Tripos contained a course on Random Variables which aimed at preparing the way for fuller statistical work in Part III.* Additional post-graduate courses were offered and, finally, to give more adequate training on the applied side, there was a Diploma course which required the post-graduate student to spend a considerable amount of his time on the application of theory in a selected field. The Diploma courses at Oxford, Manchester and Aberdeen have followed the Cambridge experiment on rather similar lines.

Approval was given for a Statistical Laboratory to be set up within the Mathematics Faculty, and after some delay due to building difficulties this was housed in a temporary building in St Andrew's Hill. When it was opened in the summer of 1949, Wishart could look back with some satisfaction at the success which he had had in grafting the teaching of statistics on to the Cambridge mathematical system. Yet recognition of his own position still came rather grudgingly, for it was not until 1953 that his appointment as Director of the Statistical Laboratory was confirmed. It began to be realized, too, as had been found elsewhere, that a statistical unit could be of better service to a university if it were free from the control of mathematics.

In his post-war published work the emphasis was again on exposition, but this was now mainly in the mathematical rather than the agricultural field. Thus between 1947 and 1951 at a time when the Institute of Actuaries was introducing additional statistical matter into its revised examination syllabus, he wrote four useful papers (**54, 55, 59, 65**) on different aspects of statistical theory for the *Journal* of the Institute's Students' Society. New contributions to theory include three *Biometrika* papers (**53, 62, 67**) dealing with his old favourite, the development of moment and cumulant theory. In the applied field, perhaps the most important contribution was a largely rewritten Part I to Wishart & Sanders's *Principles and Practice of Field Experimentation* (**75**). This admirable little book, translated into Spanish, formed the basis of his 1956 F.A.O. course in Mexico.

If his output of new scientific work was small, Wishart's last ten years were very active ones. In addition to the work of building up the Cambridge Laboratory, he was made the first Chairman of the Royal Statistical Society's Research Section, when it branched off from the Industrial and Agricultural Research Section in 1945. He also took a leading part in the Society's two post-war committees on the Teaching of Statistics. Made an Assistant Editor of *Biometrika* in 1937 and Associate Editor in 1948 he played a loyal and diligent part in maintaining the standard and tradition of this Journal.

With able colleagues available in Cambridge it was possible for him to make use of the University's newly established sabbatical year and to find other opportunities for going overseas. Thus he was in the United States for nine months in 1949, lecturing at the Universities of North Carolina and California. In the autumn of 1954 he lectured for three

* A full description of the courses available in 1955 is given as an Appendix to Wishart's contribution (**72**) to the discussion on the Teaching of Statistics held by the Royal Statistical Society in February 1955.

months at the F.A.O.'s Training Centre in Experimental Design and the Survey Technique of Experimentation held at New Delhi. Finally, after going to Mexico in January 1956 as a technical assistance expert to the Mexican Government, he was responsible for the day-to-day running of a similar F.A.O. course in Mexico City from April until his death in July.

A number of warm tributes have been paid to the value of his work on these international courses. There were many elements in his success: the carefully planned lectures, his manner of getting them across and his skill in getting back to reality after short excursions into the abstract. As a colleague of his on the Indian course remarks: 'Few of his students will forget his injunction not to substitute covariance for weed control and other aspects of good husbandry.' But apart from work of a more formal character, he had, and showed that he had, a real concern with the particular difficulties and problems of each student and colleague. His interest in students from other lands had started in his Rothamsted days and continued at Cambridge, both in the hospitality which he and his wife showed to foreign visitors in their home and through the activities of the Cambridge All Peoples Association. In both India and Mexico he was eager to learn about the art and architecture, the history and customs, of the people around him, and he learned enough Spanish to make his inaugural speech at the Central American Training Centre in this language. His friendliness and obvious enjoyment in all that was happening made a notable contribution to the success of these gatherings.

If we look back in retrospect we shall realize, I think, that John Wishart's contribution to our subject has lain not so much in any outstanding piece of mathematical research or the development of new concepts, but in the many-sided and very necessary part which he has played in the dissemination of statistical ideas and the application of statistical methods during the last 30 years. Whether it was in early trying out of the analysis of variance in the field, in convincing the agriculturalist that the principles of experimental design were not too hard after all to understand, in exploiting new techniques to the general advantage of an expanding methodology, in building up a teaching and research organization at Cambridge or in extending the activities of the Royal Statistical Society, he was there, doing his share of work to the full.

E. S. PEARSON

Bibliography of Wishart's published work

(1) Determination of $\int_0^\theta \cos^{n+1}\theta\, d\theta$ for large values of n, and its application to the probability integral of symmetrical frequency curves. (1925), *Biometrika*, **17**, 68–78.

(2) Further consideration of the integral $\int_0^\theta \cos^{n+1}\theta\, d\theta$ for large values of n. (1925), *Biometrika*, **17**, 469–72.

(3) Note on Barlow's Tables. (1926), *Nature, Lond.*, **117**, 856–7. Accurate square roots (Schlesinger, F., a reply). (1926), *Nature, Lond.*, **118**, 338–9.

(4) On Romanovsky's generalised frequency curves. (1926), *Biometrika*, **18**, 221–8.

(5) On the approximate quadrature of certain skew curves, with an account of the researches of Thomas Bayes. (1927), *Biometrika*, **19**, 1–38.

(6) On the distribution of the error of an interpolated value, and on the construction of tables (with R. A. Fisher). (1927), *Proc. Camb. Phil. Soc.* **23**, 912–21.

(7) On errors in the multiple correlation coefficient due to random sampling. (1928), *Mem. Roy. Meteor. Soc.* **2**, 29–37 and (summarized) *Quart. J. R. Met. Soc.* **54**, 308.

(8) The generalised product moment distribution in samples from a normal multivariate population. (1928), *Biometrika*, **20A**, 32–52.

(9) Table of significant values of the multiple correlation coefficient. (1928), *Quart. J. R. Met. Soc.* **54**, 258–9.
(10) Sampling errors in the theory of two factors. (1928), *Brit. J. Psychol.* **19**, 180–7.
(11) Le traitement correct des problèmes de corrélation multiple en météorologie et en agriculture. (1928), *Ass. Franc. l'Avance. des Sci., Congr. d. l. Rochelle* (pp. 4).
(12) A problem in combinatorial analysis giving the distribution of certain moment statistics. (1928), *Proc. Lond. Math. Soc.* **29**, 309–21.
(13) The correlation between product moments of any order in samples from a normal population. (1929), *Proc. Roy. Soc. Edinb.* **49**, 78–91.
(14) Fertiliser trials on the ordinary farm (with H. J. G. Hines). (1929), *J. Minist. Agric.* **36**, 524–32.
(15) A study in sampling technique: the effect of artificial fertilisers on the yield of potatoes (with A. R. Clapham). (1929), *J. Agric. Sci.* **19**, 600–18.
(16) Studies in crop variation. VII. The influence of rainfall on the yield of barley at Rothamsted (with Winifred A. Mackenzie). (1930), *J. Agric. Sci.* **20**, 417–39.
(17) The arrangement of field experiments and the statistical reduction of the results (with R. A. Fisher). (1930), *Tech. Commun. Imp. Bur. Soil Sci.* no. 10.
(18) A method of estimating the yield of a missing plot in field experimental work (with F. E. Allan). (1930), *J. Agric. Sci.* **20**, 399–406.
(19) On the secular variation of rainfall at Rothamsted. (1930), *Mem. R. Met. Soc.* **3**, 127–37.
(20) Fertiliser trials in 1929 (with H. V. Garner). (1930), *J. Minist. Agric.* **37**, 793–802.
(21) The derivation of certain high order sampling product moments from a normal population. (1930), *Biometrika*, **22**, 224–38.
(22) The analysis of variance illustrated in its application to a complex agricultural experiment on sugar beet. (1931), *Arch. Pflanzenbau*, **5**, 561–84.
(23) The mean and second moment coefficient of the multiple correlation coefficient in samples from a normal population. (1931), *Biometrika*, **22**, 353–61.
(24) The derivation of the pattern formulae of two-way partitions, from those of simpler patterns (with R. A. Fisher). (1931), *Proc. Lond. Math. Soc.* **33**, 195–208.
(25) Notes on frequency-constants. (1931), *J. Inst. Actu.* **62**, 174–7.
(26) Methods of field experimentation and the statistical analysis of the results. (1931), *Rothamsted Conferences*, XIII, 13–21.
(27) Interpolation without printed differences: Jordan's and Aitken's formulae. (1932), *Math. Gaz.* **16**, 14–25.
(28) The distribution of second order moment statistics in a normal system (with M. S. Bartlett). (1932), *Proc. Camb. Phil. Soc.* **28**, 455–9.
(29) A note on the distribution of the correlation ratio. (1932), *Biometrika*, **24**, 441–56.
(30) The generalised product moment distribution in a normal system (with M. S. Bartlett). (1933), *Proc. Camb. Phil. Soc.* **29**, 260–70.
(31) A comparison of the semi-invariants of the distributions of moment and semi-invariant estimates in samples from an infinite population. (1933), *Biometrika*, **25**, 52–60.
(32) A statistical analysis of the inter-relations of litter size and duration of pregnancy on the birth weight of rabbits (with J. Hammond). (1933), *J. Agric. Sci.* **23**, 463–72.
(33) The theory of orthogonal polynomial fitting (review). (1933), *J. R. Statist. Soc.* **96**, 487–91.
(34) Field experimentation: the modern technique. (1934), *Agric. Progr.* **11**, 149–56.
(35) Statistics in agricultural research. (1934), *J. R. Statist. Soc. Suppl.* **1**, 26–51.
(36) Bibliography of agricultural statistics, 1931–33. (1934), *J. R. Statist. Soc. Suppl.* **1**, 94–106.
(37) Analysis of variance and analysis of covariance, their meaning and their application in crop experimentation. (1934), *Report of 2nd Emp. Cotton Gr. Corp. Conf.* pp. 83–9.
(38) *Principles and Practice of Field Experimentation* (with H. G. Sanders). (1935), London, Emp. Cotton Gr. Corp.
(39) The evolution of the field experiment. (1935), *Proc. 1st Plant Breeding Conf. Nanking, China*, 1934–5, pp. 25–30.
(40) Tests of significance in analysis of covariance. (1936), *J. R. Statist. Soc. Suppl.* **3**, 79–82.
(41) Statistics in Chinese agricultural research. (1936), *J. Amer. Statist. Ass.* **31**, 127–8.
(42) A theorem concerning the distribution of joins between line segments (with H. O. Hirschfeld). (1936), *J. Lond. Math. Soc.* **11**, 227–35.
(43) The nutrition of the bacon pig. I. The influence of high levels of protein intake on growth, conformation and quality in the bacon pig (with H. E. Woodman *et al.*). (1936), *J. Agric. Sci.* **26**, 546–619.

(44) Field experiments of factorial design. (1936), *J. Agric. Sci.* **28**, 299–306.
(45) Growth-rate determinations in nutrition studies with the bacon pig, and their analysis. (1938), *Biometrika*, **30**, 16–28.
(46) Crop estimation and its relation to agricultural meteorology. (1938), *J. R. Statist. Soc. Suppl.* **5**, 20–5.
(47) Statistical treatment of animal experiments. (1939), *J. R. Statist. Soc. Suppl.* **6**, 1–22.
(48) Some aspects of the teaching of statistics. (1939), *J. R. Statist. Soc.* **102**, 532–51.
(49) *Field Trials: their Lay-out and Statistical Analysis.* (1940), Cambridge: Imperial Bureau of Plant Breeding and Genetics.
(50) *Mathematical Tables. Vol. VIII. Number-divisor Tables* (by J. W. L. Glaisher). (1940). (Editorial work in completion of Tables for the British Association Tables Committee.) Cambridge University Press.
(51) *'Student's' Collected Papers.* (1944) (edited with E. S. Pearson for the Biometrika Trustees). Cambridge University Press.
(52) Note on the probability distribution arising in the study of the Institute's examinations. (1947), *J. Inst. Actu. Stud. Soc.* **6**, 140–3.
(53) The cumulants of the z and of the logarithmic χ^2 and t distributions. (1947), *Biometrika*, **34**, 170–8.
(54) The variance ratio test in statistics. (1947), *J. Inst. Actu. Stud. Soc.* **6**, 172–84.
(55) Proof of the distributions of χ^2, of the estimate of variance, and of the variance ratio. (1947), *J. Inst. Actu. Stud. Soc.* **7**, 98–103.
(56) La estadística en Inglaterra en relación con las facultades universitarias de matematicas. (1947), Paper read to Consejo Superior de Investigaciones Cientificas, Instituto 'Jorge Juan' de Matematicas, April.
(57) Statistical aspects of demobilization in the Royal Navy. (1947), *J. R. Statist. Soc.* **110**, 27–44.
(58) Proofs of the distribution law of the second order moment statistics. (1948), *Biometrika*, **35**, 55–7.
(59) Tests of significance in the simple regression problem. (1948), *J. Inst. Actu. Stud. Soc.* **8**, 38–43.
(60) Test of homogeneity of regression coefficients, and its application in the analysis of covariance. (1948). Paper read at the Colloque de Calcul des Probabilités et de Statistique Mathématique, Lyon. (1949), *Calcul des Prob.* 93–100.
(61) The teaching of statistics. (1948), *J. R. Statist. Soc.* A, **111** (Opening contribution to the Discussion), 212–7.
(62) Cumulants of multivariate multinomial distributions. (1949), *Biometrika*, **36**, 47–58.
(63) *Analisis de la Varianza y Covarianza.* (1949), Madrid: Instituto Nacional de Investigaciones Agronomicas.
(64) *Field Trials. II. The Analysis of Covariance.* (1950), Cambridge: Commonwealth Bureau of Plant Breeding and Genetics.
(65) The variance-ratio distribution. (1951), *J. Inst. Actu. Stud. Soc.* **10**, 258–9.
(66) Biological Experimentation. (1951), *Incorp. Statistician*, **2**, 21–7.
(67) Moment coefficients of the k-statistics in samples from a finite population. (1952), *Biometrika*, **39**, 1–13.
(68) The combinatorial development of the cumulants of the k-statistics. (1952), *Trabajos de Estadistica*, **3**, 13–26.
(69) The teaching of statistics. (1952), *Incorp. Statistician*, **3**, 3–6.
(70) Orthogonal polynomial fitting (with T. Metakides). (1953), *Biometrika*, **40**, 361–9.
(71) The factorial moments of the distribution of joins between line segments. (1954), *Biometrika*, **41**, 555–6.
(72) Contribution to a discussion on the teaching of statistics at university level. (1955), *J. R. Statist. Soc.* A, **118**, 194–5, 210–11.
(73) The variance-ratio test. (1955), Int. Statist. Inst., Statistical Seminar, Rome, 1953, 9–12.
(74) Multivariate Analysis. (1955), *Applied Statistics*, **4**, 103–16.
(75) *Principles and Practice of Field Experimentation* (with H. G. Sanders) (2nd ed. 1955). Commonwealth Bureau of Plant Breeding and Genetics. (See (38) above.)
(76) 'Significant Difference' in the analysis of field trials. (1956), *Agric. Rev.* May 1956, (pp. 3).
(77) χ^2 probabilities for large numbers of degrees of freedom. (1956), *Biometrika*, **43**, 92–5.

RESTRICTED SEQUENTIAL PROCEDURES

By P. ARMITAGE

Statistical Research Unit of the Medical Research Council, London School of Hygiene and Tropical Medicine

1. Introduction

When sequential methods are advocated for industrial inspection work, the argument is usually advanced on grounds of economy. A given degree of discrimination between 'good' and 'bad' quality is achieved, by a sequential procedure, with a smaller average number of observations than that required by an equivalent non-sequential procedure. This may well be a relevant criterion when the sampling inspection is to be repeated frequently as a routine operation. In many fields of experimentation this argument carries little weight, because the uncertainty in knowing how long a particular sequential experiment will continue may outweigh any possible long-term economy. There is, however, a rather different reason why sequential procedures of some sort may be suitable for many trials in clinical and preventive medicine (which are essentially classical randomization experiments to compare different medical treatments, with human subjects as experimental units). These trials are usually sequential in the trivial sense that the subjects enter the experiment at different points of time, but this fact would not necessarily suggest sequential methods of design and analysis. The important point is that a doctor responsible for patients in a trial will frequently regard it as unethical to continue the trial if he is convinced that a difference between the effects of two treatments has appeared, because this would mean withholding from a patient under his care a treatment which he regarded as better than that being given. This consideration will be most cogent when life and death are involved; it may not apply at all if the illnesses being treated are fairly trivial, but it will usually be present to some extent. There is, therefore, a natural tendency for the organizers of a clinical or preventive trial to examine the observations at various points of time, and to stop or modify the trial if a difference between treatment effects is apparent. Sequential experimentation of this sort, however reasonable, will to some extent affect the validity of any probability statements involving integration over sample space (Anscombe, 1954), and it is natural to investigate the possibility of carrying out the trial according to some well-defined sequential procedure for which valid probability statements can be made.

In an earlier paper (Armitage, 1954) I suggested that comparisons of two treatments could be made by fairly straightforward modifications of Wald's methods. Means could be compared by two-sided sequential t-tests (National Bureau of Standards, 1951), and proportions could be compared by a two-sided analogue of Wald's one-sided test for comparative trials. (The test for proportions has been proposed also by de Boer (1953).) These procedures retain the feature, common to all probability-ratio sequential tests, that the number of observations required before a boundary is reached is unlimited. This property is in practice unattractive, and although a Wald sequential procedure can be truncated the effects of truncation are not generally known. To preserve, even approximately, the nominal 'risks of error' the truncation must be applied after a fairly large number of observations, and the distributions of the sample number still have very high variability (for examples of such distributions, see Baker, 1950). Such variability discourages the use of this type of

procedure for medical trials, since it is clearly inconvenient to plan a trial the duration of which has an *expectation* of between, say, 2 and 9 months (depending on the actual difference between the effects of the treatments), but which may in an individual instance have to be continued for as long as 3 years.

The type of procedure which seems to be required is one which incorporates truncation as an integral part of the scheme, which allows much less variability of sample number than do Wald schemes, and for which some at least of the probabilistic properties are known. Bross (1952) constructed some closed sequential designs for the comparison of two proportions, but although they satisfy the criteria stated above to be desirable, each design had to be constructed specially, to some extent by trial and error, and only two have so far been published. In the present paper I consider a class of closed sequential procedures, which I have called 'restricted sequential procedures'. These seem to satisfy the above requirements, but their probabilistic properties depend on a diffusion approximation, the adequacy of which has not been fully investigated. The main feature of the method is that the sample number cannot be greater than some predetermined value, N. As sampling proceeds, a sample path may be drawn on a diagram, and sampling will stop after less than N observations if one of two 'outer boundaries' is crossed. These outer boundaries are determined by the requirement that the likelihood ratio of either of two alternative hypotheses, to the null hypothesis, takes some preassigned high value. It may be arranged that the probability, on the null hypothesis, that the sample path will stop on an outer boundary takes a predetermined small value, 2α. If the procedure is used to provide a two-sided test of the null hypothesis, sample paths reaching the outer boundaries will then correspond to results significant at the 2α level; those failing to reach the outer boundaries before N observations will correspond to non-significant results. The greater the apparent departure from the null hypothesis, the sooner will one of the outer boundaries be hit, and sampling brought to a close—a feature which seems to accord with medical requirements.

These procedures were developed with medical applications in mind, but they may be of some use in industrial or other situations. I have deliberately not referred to them exclusively as sequential *tests*, because the point of view adopted here is that a sequential procedure is primarily the specification of a stopping rule. If its probabilistic properties are sufficiently well known, a given stopping rule may be used to different ends: it may provide an acceptance procedure, in which case the appropriate formulation may be in terms of risks of acceptance or rejection; it may provide a test of a null hypothesis, in which case its power function may be of interest; it may be used for the estimation of an unknown parameter; or it may be used in more than one of these ways on the same data.

For an account of a clinical trial conducted by these methods, see Snell & Armitage (1957).

2. Normal distribution with known variance

2·1. Suppose that observations $x_i (i = 1, 2, \ldots)$ are drawn at random from a normal distribution with unknown mean μ and known variance σ^2. Let $y_n = \sum_{i=1}^{n} x_i$. 'Restricted sequential procedures' will consist in sampling until one of the three following boundaries is reached:

(i) the 'upper boundary', U: $y_n = a + bn$ $(a > 0)$;

(ii) the 'lower boundary', L: $y_n = -a - bn$ $(a > 0)$;

(iii) the 'middle boundary', M: $n = N$.

U and L will be called 'outer boundaries'. In order that neither outer boundary crosses $y_n = 0$ for $n < N$ we shall require $a + bN \geqslant 0$.

Suppose first that only U and M are present. If the discrete steps in n are replaced by continuous movement in time, where the time unit corresponds to a single observation, the random walk may be approximated by the one-dimensional diffusion process with drift $\mu - b$ per unit time, growth in variance at a rate σ^2, and an absorbing barrier at a. Hence, the probability of crossing U with not more than N observations is approximated to by the probability of absorption before time N, which is given (Bartlett, 1946) by

$$P_0(\mu, N) = 1 - F\left\{\frac{a-(\mu-b)N}{\sigma\sqrt{N}}\right\} + \exp\left\{\frac{2a(\mu-b)}{\sigma^2}\right\} F\left\{\frac{-a-(\mu-b)N}{\sigma\sqrt{N}}\right\}, \tag{1}$$

where $F(u) = \int_{-\infty}^{u} (2\pi)^{-\frac{1}{2}} e^{-\frac{1}{2}t^2} dt$. This formula is valid for $a > 0$, as is appropriate here. The same approach to this problem is suggested by Page (1957).

If L is now introduced, the probability of absorption on U is affected. The expression (1) should be reduced by the probability that the path of the diffusing particle crosses first L and then U before $n = N$. This effect appears to be less important with the convergent or divergent boundaries considered here than it is with the Wald type of parallel-line boundaries.

For divergent boundaries (when $b > 0$), the probability that a path, starting at a point on L, will cross U before $n = N$, is less than it would be for a path starting at the point $(n = 0, y_n = -a)$. For such a path the probability of crossing the line $y_n = bn$ before $n = N$ is equal to $P_0(\mu, N)$, and the conditional probability of then crossing U before $n = N$ is less than $P_0(\mu, N)$. The total probability is therefore less than $\{P_0(\mu, N)\}^2$. But the probability that a path starting at the origin will cross L before $n = N$ is less than $P_0(-\mu, N)$. Hence, an upper bound to the probability that a path crosses L and then U before $n = N$ is $P_0(-\mu, N)\{P_0(\mu, N)\}^2$, and an upper bound to the proportionate error (i.e. the ratio of the error to the nominal value $P_0(\mu, N)$) is $P_0(-\mu, N) P_0(\mu, N)$. For particular choices of a, b and N this upper bound may not be very small, but in the applications discussed below it will be shown to be small.

No analogous upper bound has been obtained for convergent boundaries, and in most of the following discussion we therefore assume that $b > 0$.

2·2. The specification of the boundaries requires three arbitrary constants: a, b and N. In general, therefore, a set of boundaries will not be determined uniquely by the requirement that the procedure should satisfy only one or two independent conditions. Consider first the problem of determining values of a, b and N satisfying the following two conditions:

(*a*) On the hypothesis H_0 that $\mu = 0$, the probabilities that sampling will end on each of the outer boundaries are $\alpha < \frac{1}{2}$ (these probabilities clearly being equal by symmetry). If the procedure is used to provide a significance test of H_0, a sample path ending on one of the outer boundaries may then be regarded as significant at a level 2α.

(*b*) On the hypothesis H_+ that $\mu = \mu_1 > 0$, the probability of reaching U is $1 - \beta$ $(> \alpha)$. (By symmetry, $1 - \beta$ will also be the probability of reaching L on the hypothesis H_- that $\mu = -\mu_1$.)

Since only two restrictions have been imposed on the three constants a, b and N, we should expect them to determine a family of restricted sequential designs. The approximate solution below provides one design out of this family.

Choose
$$a = \frac{\sigma^2}{\mu_1} \ln \frac{1-\beta}{\alpha}, \quad b = \frac{\mu_1}{2}. \tag{2}$$

Then, on U, the likelihood ratio of H_+ to H_0 is $(1-\beta)/\alpha$. Whether L is present or not, the probabilities, on H_+ and H_0, of reaching any specified set of points on U will be in the ratio $1-\beta$ to α (neglecting the overshooting of the boundaries). In particular, if $P(\mu, N)$ is the probability of absorption on U in less than N observations, no other boundary having previously been reached, and we choose N such that $P(\mu_1, N) = 1-\beta$, it will follow that (again neglecting the overshooting of the boundaries) $P(0, N) = \alpha$. Conditions (a) and (b) will then be satisfied. Approximating to $P(\mu, N)$ by $P_0(\mu, N)$, given by (1), we require

$$\beta = F\left(\frac{\ln\{(1-\beta)/\alpha\}}{\Delta\sqrt{N}} - \frac{\Delta\sqrt{N}}{2}\right) - \left(\frac{1-\beta}{\alpha}\right) F\left(-\frac{\ln\{(1-\beta)/\alpha\}}{\Delta\sqrt{N}} - \frac{\Delta\sqrt{N}}{2}\right), \tag{3}$$

where $\Delta = \mu_1/\sigma$. Since α, β and Δ are known, (3) may be solved for N, by successive approximation. In fact, for given α and β, (3) may be solved for $\Delta\sqrt{N}$; the solutions, $\Delta\sqrt{N_1}$, for various combinations of α and β are shown in Table 1.

Table 1. *Normal distribution with known variance: values of* $\Delta\sqrt{N_1}$, *for various values of* α *and* β

β \ α	0·050	0·025	0·005
0·10	3·41	3·72	4·32
0·05	3·92	4·22	4·80
0·01	4·95	5·23	5·79

The use of the single-boundary diffusion approximation (1) is justified since the upper bound to the proportionate error in α (and hence in $1-\beta$) is $\{P_0(0, N)\}^2 = \alpha^2$. In practice α will usually be chosen to be fairly small (say 0·05 or less), since in the practical applications envisaged at the moment α may be regarded as the (one-sided) significance level at which one wishes to use less than the maximum sample number, N. The other approximation involved—that of ignoring the overshooting of the boundaries—is of the same type as that involved in the formulae customarily used in Wald's sequential tests. Wald's formulae also can be obtained by diffusion theory (Bartlett, 1946), and it would seem a reasonable conjecture that the accuracy of the approximation in the present schemes would be similar to that of Wald's approximations.

One other member of the family of restricted sequential procedures satisfying (a) and (b) is the degenerate case obtained by putting $a = \infty, b = -\infty, N = N_0$ and $a + bN_0 = u_\alpha \sigma\sqrt{N_0}$, where

$$F(u_\alpha) = 1-\alpha, \quad F(u_\beta) = 1-\beta,$$

and
$$\Delta\sqrt{N_0} = u_\alpha + u_\beta.$$

This is the fixed-sample-size procedure satisfying (a) and (b). Since the procedure with $N = N_1$, and the fixed-sample-size procedure with $N = N_0$ both satisfy (a) and (b), and since the former permits sampling to stop for $n < N_1$, we should expect to find $N_1 > N_0$. The ratios N_1/N_0 are given in Table 2 for various combinations of α and β.

Other members of the family of procedures satisfying (a) and (b) could be found by trial and error, but no attempt to do this has yet been made.

Table 2. *Normal distribution with known variance: values of N_1/N_0, for various values of α and β*

β \ α	0·050	0·025	0·005
0·10	1·355	1·313	1·252
0·05	1·417	1·368	1·295
0·01	1·553	1·490	1·393

2·3. It might in practice be more convenient to specify in advance the maximum sample number N, rather than the probability β. We therefore consider the problem of determining values of a and b satisfying condition (a) as before, given that the maximum sample number is N. Again, we should expect a family of procedures satisfying these conditions.

Since α and N are known, equation (3) may be solved for Δ by successive approximation for any chosen value of β. Denote the solution by $\Delta(\beta)$. Each pair of values of $(\beta, \Delta(\beta))$ determines the value of a and b. A family of procedures may therefore be generated by allowing β to assume different values. We shall here relax the condition required in § 2·2, that $\beta < 1-\alpha$. This is a natural requirement to impose if the boundaries are to be determined by considerations of power, but it is instructive to consider the wider class of boundaries obtained by allowing β to vary between 0 and 1. In § 2·4 we prove that $\Delta(\beta)$ is a single-valued monotonic decreasing function of β, ranging from $+\infty$ to $-\infty$ as β ranges from 0 to 1. For $\beta \lesseqgtr 1-\alpha$, $\Delta(\beta) \gtreqless 0$. Hence, for all β, the solutions satisfy

$$a\left(=\frac{\sigma}{\Delta(\beta)}\ln\frac{1-\beta}{\alpha}\right)>0,$$

as required. (This result is intuitively plausible, since $(1-\beta)/\alpha$ is the likelihood ratio, at points on U, of the hypothesis that $\mu = \Delta(\beta)\sigma$ to the hypothesis that $\mu = 0$. It is therefore reasonable that $(1-\beta)/\alpha > 1$ if $\Delta(\beta) > 0$ and $(1-\beta)/\alpha < 1$ if $\Delta(\beta) < 0$.) It is also proved in § 2·4 that $a + bN > u_\alpha > 0$; the requirement of § 2·1 that $a + bN > 0$ is thus satisfied. For $\beta > 1-\alpha$, $b\,(= \sigma\Delta(\beta)/2) < 0$ and the boundaries are convergent. Since, for convergent boundaries, no upper bound is available for the proportionate error involved in using the single-boundary diffusion theory, the procedures obtained for values of $\beta > 1-\alpha$ are at present of less practical interest than those for $\beta < 1-\alpha$. Table 3 gives values of a/σ, b/σ and $\Delta(\beta)$ for various values of β, in the particular case where $\alpha = 0{\cdot}025$, $N = 50$. These values were obtained by successive approximation. The corresponding upper boundaries are illustrated in Fig. 1.

The limiting forms as $\beta \to 0$ and $\Delta(\beta) \to \infty$ are of little practical interest, because when the slope of the outer boundary is very steep the total probability of hitting the upper boundary, α, is largely concentrated at very low values of n, and in these circumstances the diffusion process (which allows the boundary to be crossed for any $n > 0$) is a poor approximation to a process involving only integral n.

The horizontal boundary obtained for $\beta = 1-\alpha$ is of some interest, as procedures with one horizontal boundary have been discussed by Rao (1950). The value of a is indeterminate, but the limiting value, a_0, is shown in § 2·4 to satisfy the equation

$$F(a_0/\sigma\sqrt{N}) = 1-\tfrac{1}{2}\alpha. \tag{4}$$

Table 3. *Normal distribution with known variance: values of a/σ, b/σ and $\Delta(\beta)$ for restricted sequential procedures with $\alpha = 0{\cdot}025$, $N = 50$, and various values of β (cf. Fig. 1)*

β	a/σ	b/σ	$\Delta(\beta)$
0·001	4·02	0·458	0·916
0·010	4·97	0·370	0·740
0·050	6·10	0·298	0·596
0·100	6·82	0·263	0·525
0·500	9·95	0·150	0·301
0·900	13·7	0·051	0·101
0·975	15·8	0·000	0·000
0·980	16·1	−0·007	−0·014
0·990	17·0	−0·027	−0·054
0·999	19·5	−0·082	−0·165

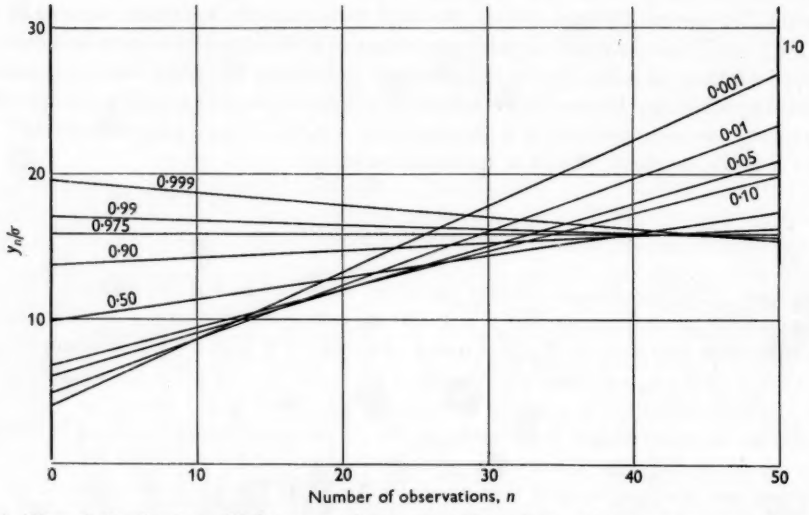


Fig. 1. Normal distribution with known variance: upper boundaries, U, for various procedures with $\alpha = 0{\cdot}025$, $N = 50$. The boundaries are generated by the parameter β, the values of which are shown in the diagram. For $\beta = 1{\cdot}0$, the outer boundary is vertical, corresponding to a fixed-sample-size procedure.

In the notation of § 2·2, $a_0 = u_{\frac{1}{2}\alpha}\sigma\sqrt{N}$. The boundary therefore meets $n = N$ at a point corresponding to a fixed-sample-size significance level of $\frac{1}{2}\alpha$, on a one-sided test. The upper bound to the proportionate error, derived for divergent boundaries, is readily seen to be valid also for horizontal boundaries. Moreover, an exact expression is available for the

distribution of absorption time in the diffusion problem with two parallel boundaries (e.g. Feller, 1950, p. 304, Problem 7). In this particular case the upper bound for the proportionate error appears to be unnecessarily wide. For instance, if the boundaries are set at $y_n = \pm a_0$, where a_0 is given by (4) with $\alpha = 0{\cdot}025$, the probability, on H_0, of absorption before $n = N$ falls short of 0·05 by less than 10^{-13}.

Consider now the effect of replacing the diffusion process by discrete sampling, in the situation considered by Rao. Then, if U is horizontal and L is omitted (as in Rao's procedure), a_0 (defined by (4)) provides a rather close upper bound to the exact value A_0 required to satisfy given values of α and N. For, if the equation of U is $y_n = a_0$, any sample path which, if continued to $n = N$ satisfies $y_N > a_0$, must cross U at some $n \leqslant N$. Consider a path which first crosses U at $n = n_0 < N$. On hypothesis H_0, the probability that such a path, when continued to $n = N$, will satisfy $y_N > a_0$, is slightly greater than $\frac{1}{2}$, for all n_0. (It would equal $\frac{1}{2}$ if the paths did not overshoot the boundary.) If the probability, on H_0, of first crossing U at $n = n_0$ is $\alpha(n_0)$, then the probability that $y_N > a_0$ is slightly greater than

$$\frac{1}{2}\sum_{n_0=1}^{N-1}\alpha(n_0) + \alpha(N),$$

which is slightly greater than

$$\frac{1}{2}\sum_{1}^{N}\alpha(n_0).$$

But, by the definition of a_0, the probability that $y_N > a_0$ is $\frac{1}{2}\alpha$. Hence

$$\alpha > \sum_{1}^{N}\alpha(n_0),$$

and, in order to replace the inequality by an equality, a_0 must be reduced slightly. Hence a_0 provides a fairly close upper limit to A_0, rather better than that given by Rao (1950), as may be seen from Table 4.

Table 4. *Normal distribution with known variance: procedure with single boundary at $y_n = A_0$, restricted at $n = N$. Upper bounds for $A_0/\sigma\sqrt{N}$, for various values of α*

α	Upper bound given by	
	$u_{\frac{1}{2}\alpha} = a_0/\sigma\sqrt{N}$	Rao, p. 365
0·050	1·9600	2·0626
0·025	2·2414	2·3376
0·010	2·5758	2·6655

The other limiting form, obtained as $\Delta(\beta) \to -\infty$ and $\beta \to 1$, is of course a fixed-sample-size procedure, with the upper boundary consisting of those points significant at a probability level α, on a one-sided test. This is proved in § 2·4. It may be noted that in this case the proportionate error introduced by using the one-boundary diffusion approximation is zero, even though $b < 0$.

The choice of a restricted sequential procedure, for given values of α and N, is almost embarrassingly wide, and seems to depend on how 'sequential' we wish the procedure to

be—to what extent, that is, we wish the sample number to be reduced below the maximum of N, as the mean, $\bar{x}$, increases. At one extreme we have the fixed-sample-size procedure, in which the sample number is always N. At the other extreme we have procedures, based on low values of β and high values of $\Delta(\beta)$, which terminate very much earlier for fairly high values of $\bar{x}$. It seems likely (but has not been proved) that, for given α and N, the procedures based on the higher values of β (i.e. the less 'sequential' ones) are the more powerful, in the sense that for a given value of $\mu > 0$ the probability of crossing the upper boundary is greater. In the absence of any exact formulation of risk functions, it would seem reasonable (judging from Fig. 1) to use procedures based on values of β between 0·001 and 0·50.

If the given value of α is one of those used in Table 1 (0·005, 0·025 or 0·05), the boundaries of three procedures (based on $\beta = 0{\cdot}01$, 0·05 and 0·10) are easily found, for division of the tabulated value by $\sqrt{N}$ gives $\Delta(\beta)$, and, from (2),

$$a = \frac{\sigma}{\Delta(\beta)} \ln \frac{1-\beta}{\alpha}, \quad b = \frac{\sigma \Delta(\beta)}{2}.$$

2·4. We now prove the results stated in § 2·3. We have first

THEOREM 1. *For given values of α and N, the solution $\Delta(\beta)$ of* (3) *is a single-valued function of β, decreasing monotonically from $+\infty$ to $-\infty$ as β increases from* 0 *to* 1.

Consider the function

$$\phi = F\left(\frac{\ln\{(1-\beta)/\alpha\}}{\Delta\sqrt{N}} - \frac{\Delta\sqrt{N}}{2}\right) - \left(\frac{1-\beta}{\alpha}\right) F\left(-\frac{\ln\{(1-\beta)/\alpha\}}{\Delta\sqrt{N}} - \frac{\Delta\sqrt{N}}{2}\right) - \beta. \tag{5}$$

(3) is satisfied if and only if $\phi = 0$.

From (5), after some reduction

$$\frac{\partial\phi}{\partial\Delta} = -\left\{\frac{2(1-\beta)}{\pi\alpha N}\right\}^{\frac{1}{2}} \frac{\ln\left(\frac{1-\beta}{\alpha}\right)}{\Delta^2} \exp - \frac{1}{2}\left\{\frac{\left(\ln\frac{1-\beta}{\alpha}\right)^2}{\Delta^2 N} + \frac{\Delta^2 N}{4}\right\}$$

$$\lessgtr 0 \text{ according as } \beta \lessgtr 1-\alpha.$$

As a function of Δ, ϕ has a discontinuity at $\Delta = 0$. Consideration of the limiting values of ϕ as $\Delta \to \pm\infty$ and ± 0, and of the sign of $\partial\phi/\partial\Delta$, shows that $\phi(\Delta)$ assumes one of two distinct forms, according as $\beta < 1-\alpha$ or $\beta > 1-\alpha$.

For $\beta < 1-\alpha$, ϕ decreases from $(1-\beta)(1-1/\alpha)$ to $(1-\beta)(1-1/\alpha)-1$ as Δ increases from $-\infty$ to -0, and from $1-\beta$ to $-\beta$ as Δ increases from $+0$ to $+\infty$.

For $\beta > 1-\alpha$, ϕ increases from $(1-\beta)(1-1/\alpha)$ to $1-\beta$ as Δ increases from $-\infty$ to -0, and from $(1-\beta)(1-1/\alpha)-1$ to $-\beta$ as Δ increases from $+0$ to $+\infty$.

The solution $\Delta(\beta)$ of (3) must therefore be a single-valued function of β, satisfying

$$\Delta(\beta) \gtrless 0 \quad \text{as} \quad \beta \lessgtr 1-\alpha.$$

To prove the monotonicity of $\Delta(\beta)$, consider $\partial\phi/\partial\beta$. We find, after some reduction,

$$\frac{\partial\phi}{\partial\beta} = -\frac{1}{\Delta}\left\{\frac{2}{\pi\alpha(1-\beta)N}\right\}^{\frac{1}{2}} \exp - \frac{1}{2}\left\{\frac{\left(\ln\frac{1-\beta}{\alpha}\right)^2}{\Delta^2 N} + \frac{\Delta^2 N}{4}\right\} + \frac{1}{\alpha} F\left(-\frac{\ln\frac{1-\beta}{\alpha}}{\Delta N^{\frac{1}{2}}} - \frac{\Delta N^{\frac{1}{2}}}{2}\right) - 1. \tag{6}$$

For $\Delta > 0$, $\beta < 1-\alpha$,

$$u \equiv -\frac{\ln\frac{1-\beta}{\alpha}}{\Delta N^{\frac{1}{2}}} - \frac{\Delta N^{\frac{1}{2}}}{2} < 0.$$

Hence, using the inequality on Mills's ratio given by Gordon (1941),

$$F(u) < -u^{-1}(2\pi)^{-\frac{1}{2}} e^{-\frac{1}{2}u^2}$$
$$< (2\Delta^{-1}N^{-\frac{1}{2}})(2\pi)^{-\frac{1}{2}} e^{-\frac{1}{2}u^2}$$
$$< \frac{1}{\Delta}\left(\frac{2\alpha}{\pi(1-\beta)N}\right)^{\frac{1}{2}} \exp -\frac{1}{2}\left\{\frac{\left(\ln\frac{1-\beta}{\alpha}\right)^2}{\Delta^2 N} + \frac{\Delta^2 N}{4}\right\}.$$

Hence $\partial\phi/\partial\beta < -1$, and from previous remarks about the behaviour of ϕ it immediately follows that for $\beta < 1-\alpha$ and $\Delta(\beta) > 0$, $\Delta(\beta)$ is a monotonically decreasing function of β, the limit of which, as $\beta \to 0$, is $+\infty$.

For $\Delta < 0$, $\beta > 1-\alpha$, $\partial\phi/\partial\beta$ cannot have the same sign for all Δ, since the upper and lower limits, $1-\beta$ and $(1-\beta)(1-1/\alpha)$, converge as $\beta \to 1$. But, for $\phi \leqslant 0$, by (5),

$$\frac{1}{\alpha}F\left(-\frac{\ln\frac{1-\beta}{\alpha}}{\Delta N^{\frac{1}{2}}} - \frac{\Delta N^{\frac{1}{2}}}{2}\right) \geqslant \frac{1}{1-\beta}F(v) - \frac{\beta}{1-\beta}, \tag{7}$$

where

$$v \equiv \frac{\ln\frac{1-\beta}{\alpha}}{\Delta N^{\frac{1}{2}}} - \frac{\Delta N^{\frac{1}{2}}}{2} > 0,$$

and, by Gordon's inequality,

$$F(v) = 1 - F(-v) > 1 - v^{-1}(2\pi)^{-\frac{1}{2}} e^{-\frac{1}{2}v^2}$$
$$> 1 + \frac{1}{\Delta}\left(\frac{2(1-\beta)}{\pi\alpha N}\right)^{\frac{1}{2}} \exp -\frac{1}{2}\left\{\frac{\left(\ln\frac{1-\beta}{\alpha}\right)^2}{\Delta^2 N} + \frac{\Delta^2 N}{4}\right\}.$$

Hence, the left-hand side of (7)

$$> 1 + \frac{1}{\Delta}\left(\frac{2}{\pi\alpha(1-\beta)N}\right)^{\frac{1}{2}} \exp -\frac{1}{2}\left\{\frac{\left(\ln\frac{1-\beta}{\alpha}\right)^2}{\Delta^2 N} + \frac{\Delta^2 N}{4}\right\},$$

and, from (6),

$$\partial\phi/\partial\beta > 0.$$

Hence, and from previous remarks about the behaviour of ϕ, it follows that for $\beta > 1-\alpha$ and $\Delta(\beta) < 0$, $\Delta(\beta)$ is a monotonically decreasing function of β, approaching the limiting value of $-\infty$ as $\beta \to 1$.

This completes Theorem 1. We have not strictly proved, as part of the theorem, that $\Delta(\beta) \to 0$ as $\beta \to 1-\alpha$. It is easily seen, writing $1-\beta = \alpha(1+\delta)$ in (3) and letting $\delta \to 0$, that a solution of (3) is possible only if $\Delta(\beta) = O(\delta)$.

To Theorem 1 we have the following

COROLLARY. *Given N, any restricted sequential procedure satisfying $P_0(0, N) = \alpha$* (i.e. *having a specified risk according to the diffusion approximation*) *is one of the family generated by the variable β.*

For, the slope b of the boundaries determines a value of $\Delta(\beta)$, and hence of β, and the corresponding boundaries must coincide with those given, since a must be a single-valued function of $P_0(0, N)$.

The remaining results stated in § 2·3 are included in

THEOREM 2. *The ordinate, $a+bN$, at the point of intersection of U and M, has a lower limit of $u_\alpha\sigma\sqrt{N}$, approached as $\beta\to 1$. As $\beta\to 1-\alpha$, $a+bN\to u_{\frac{1}{2}\alpha}\sigma\sqrt{N}$.*

From (2),
$$\frac{a+bN}{\sigma\sqrt{N}}=\frac{\ln\dfrac{1-\beta}{\alpha}}{\Delta(\beta)\sqrt{N}}+\frac{\Delta(\beta)\sqrt{N}}{2},$$

and, from (3),
$$1-F\left(\frac{a+bN}{\sigma\sqrt{N}}\right)=F\left(\frac{-a-bN}{\sigma\sqrt{N}}\right)$$
$$=\frac{\alpha}{1-\beta}\left\{F\left(\frac{\ln\dfrac{1-\beta}{\alpha}}{\Delta(\beta)\sqrt{N}}-\frac{\Delta(\beta)\sqrt{N}}{2}\right)-\beta\right\}$$
$$<\alpha.$$

Hence $(a+bN)/\sigma\sqrt{N}$ has a lower bound of u_α, and tends to u_α as
$$F\left(\frac{\ln\dfrac{1-\beta}{\alpha}}{\Delta(\beta)\sqrt{N}}-\frac{\Delta(\beta)\sqrt{N}}{2}\right)\to 1,$$
i.e. as $\beta\to 1$, and $\Delta(\beta)\to-\infty$.

The second part of the theorem follows from the result indicated at the end of Theorem 1, that as $\beta\to 1-\alpha$, $\Delta(\beta)=O(\delta)$, where $1-\beta=\alpha(1+\delta)$. Then $(a+bN)/\sigma\sqrt{N}\to\delta/\Delta(\beta)\sqrt{N}$, and (3) becomes
$$1-\alpha\sim F(\delta/\Delta(\beta)\sqrt{N})-(1+\delta)\,F(-\delta/\Delta(\beta)\sqrt{N})$$
$$\sim 2F(\delta/\Delta(\beta)\sqrt{N})-1.$$

Hence $F(\delta/\Delta(\beta)\sqrt{N})\sim 1-\frac{1}{2}\alpha$, and the result follows.

3. COMPARATIVE BINOMIAL TRIALS

3·1. Wald (1947) proposed a sequential method for testing the difference between the means of two binomial distributions, the non-sequential analogue of which is asymptotically fully efficient. Suppose that individuals are to be submitted to one of two treatments, T_1 and T_2, with 'success-rates' π_1 and π_2 respectively. Pairs of individuals are allocated randomly to T_1 and T_2, and the failure or success of each member of the pair is recorded. A pair showing success on T_1 and failure on T_2 will be denoted by SF, and so on. Following Bross (1952), we shall use the term 'tied pairs' for those recorded as SS or FF, and 'untied pairs' for those recorded as FS or SF. Wald's method is based entirely on the untied pairs. Let θ be the probability that an untied pair is an FS. Then
$$\theta=\frac{(1-\pi_1)\,\pi_2}{(1-\pi_1)\,\pi_2+\pi_1(1-\pi_2)}.$$

$\theta\gtreqless\frac{1}{2}$ according as $\pi_2\gtreqless\pi_1$, and the problem of testing the difference between the two binomial parameters π_1 and π_2 reduces to that of testing the departure of the single binomial parameter θ from the value $\frac{1}{2}$.

We follow here the same basic approach to the problem of comparing two binomial parameters π_1 and π_2, and consider restricted sequential procedures for sampling the population of untied pairs, of which proportions θ and $1-\theta$ are of types FS and SF respectively.

3·2. Suppose that, of the first n untied pairs, n_1 are of type FS and n_2 of type SF. Let $y_n = 2n_1 - n = n_1 - n_2$. By analogy with the problem treated in § 2·2, we shall first consider the specification of upper and lower boundaries U and L, and a middle boundary M with equation $n = N$, such that

(a) on the hypothesis H_0 that $\theta = \frac{1}{2}$, the probabilities that sampling will end on each of the outer boundaries are $\alpha < \frac{1}{2}$;

(b) on the hypothesis $H_+[H_-]$ that $\theta = \theta_1 > \frac{1}{2}$ $[\theta = 1 - \theta_1 < \frac{1}{2}]$, the probability of reaching $U[L]$ is $1 - \beta$ $(> \alpha)$. The logarithm of the likelihood ratio of H_+ to H_0 is

$$l = \tfrac{1}{2} y_n \log\{\theta_1/(1-\theta_1)\} - n \log\{\tfrac{1}{2}\theta_1^{-\frac{1}{2}}(1-\theta_1)^{-\frac{1}{2}}\}.$$

As in § 2·2 we arrange that, on U, $l = \log\{(1-\beta)/\alpha\}$. This gives the equations to U and L as $y_n = a + bn$ and $y_n = -a - bn$, respectively, where

$$a = \frac{2\log\{(1-\beta)/\alpha\}}{\log\{\theta_1/(1-\theta_1)\}} \tag{8}$$

and

$$b = \frac{2\log\{\tfrac{1}{2}\theta_1^{-\frac{1}{2}}(1-\theta_1)^{-\frac{1}{2}}\}}{\log\{\theta_1/(1-\theta_1)\}}.$$

The requirements (a) and (b) will both be satisfied (apart from neglect of the overshoot of the boundaries) if we choose N to satisfy (b).

Now, for sequential sampling from a binomial population with any boundaries, the probability of reaching a particular boundary point can be calculated exactly by enumeration of the number of paths from the origin, reaching that point without previously crossing a boundary. The required value of N could therefore be obtained by straightforward, although laborious, calculation of the probabilities, on H_+, of reaching the various boundary points on U. It will be useful, however, to have a manageable approximation to the value of N, and this can be found by applying to the binomial random walk the diffusion approximation already used in § 2.

We have $y_n = \sum_{i=1}^{n} x_i$, where the x_i are independent variates taking values $+1$ and -1 with probabilities θ and $1-\theta$ respectively. The random walk will therefore be approximated by a diffusion process with drift $\mu - b$ per unit time, growth in variance at a rate σ^2, and an absorbing barrier at a, where

$$\mu = 2\theta - 1 \quad \text{and} \quad \sigma^2 = 4\theta(1-\theta).$$

Requirement (b) will be approximately satisfied then, if (from (1) and the subsequent discussion) we ensure that

$$\beta = F\left(\frac{a - m_1 N}{\sigma_1\sqrt{N}}\right) - \exp\left(\frac{2am_1}{\sigma_1^2}\right) F\left(\frac{-a - m_1 N}{\sigma_1\sqrt{N}}\right), \tag{9}$$

where a is given by (8), $\qquad \sigma_1^2 = 4\theta_1(1-\theta_1)$,

and

$$m_1 = 2\theta_1 - 1 - \frac{2\ln(1/\sigma_1)}{\ln\{\theta_1/(1-\theta_1)\}}.$$

Table 5 gives the solution $N = N_1$ of (9), and the values of a and b, for $\alpha = 0{\cdot}025$, $\beta = 0{\cdot}05$ and various values of θ_1.

As in the corresponding problem considered in § 2·2, the solution tabulated in Table 5 provides only one member of a family of procedures approximately satisfying (*a*) and (*b*). (The discreteness of the variate x prevents an *exact* solution being found in general.) Another member of the family is a fixed-sample-size procedure with $N = N_0$, say. The choice amongst neighbouring values of N_0 each providing approximations to α and β may not be clear, but in practice suitable values may be obtained from the following normal approximation:

$$N_0 = \{\theta_1^{\frac{1}{2}}(1-\theta_1)^{\frac{1}{2}}u_\beta + \tfrac{1}{2}u_\alpha\}^2/(\theta_1 - \tfrac{1}{2})^2,$$

where, as in § 2·2, $F(u_\alpha) = 1-\alpha$, $F(u_\beta) = 1-\beta$.

Values of N_0 are given in Table 5. Values of N_1/N_0 are shown in the last column of Table 5; for θ_1 near 0·5, the ratio is close to the normal equivalent 1·368 given in Table 2, as might be expected.

Table 5. *Procedure for comparison of two proportions. Parameters of boundaries; maximum number of untied pairs, N_1; with nearest boundary point at N_1' and equivalent fixed-sample-size, N_0; for $\alpha = 0{\cdot}025$, $\beta = 0{\cdot}05$ and various values of θ_1*

θ_1	Parameters of boundaries		Maximum number of untied pairs		Equivalent fixed-sample-size, N_0	N_1/N_0
	a	b	N_1	N_1'		
0·55	36·25	0·0501	1778	1778	1294	1·37
0·60	17·94	0·1007	438	439	319	1·37
0·65	11·75	0·1524	192	191	138	1·39
0·70	8·59	0·2058	105	104	75	1·40
0·75	6·62	0·2619	65	66	46	1·41
0·80	5·25	0·3219	43	44	30	1·43
0·85	4·19	0·3882	30	30	20	1·50
0·90	3·31	0·4650	21	22	14	1·50
0·95	2·47	0·5640	14	15	9	1·56

The outer boundaries U and L can be crossed only at a discrete set $\{\mathcal{N}\}$ of values of n, and it would seem reasonable to place M at $n = N_1'$, where N_1' is the member of $\{\mathcal{N}\}$ nearest to N_1. Values of N_1' are given in Table 5.

A second modification is to replace M by the wedge-shaped boundary M' illustrated in Fig. 2. Any path crossing M' must also, if continued, cross M rather than U or L. Hence, the replacement of M by M' does not affect the probabilities of reaching any boundary point on U or L, whereas the average sample number for any θ must be reduced.

3·3. As in the similar problem considered in § 2·3, we might wish to specify in advance the maximum number of untied pairs, N, in addition to the probability α. As in § 2·3, we could generate a family of procedures satisfying these requirements, by solving (9) for θ_1, for any chosen value of β. This has not been carried out, but for $\alpha = 0{\cdot}025$ one member of

the family (corresponding to $\beta = 0{\cdot}05$) can be obtained from Table 5, by interpolating for the specified value of N in the column headed N_1.

It would perhaps be possible to prove results analogous to those in § 2·4, but such results have not yet been obtained.

3·4. The theoretical basis proposed above, for restricted sequential procedures to compare binomial variates, involves a number of approximations: the neglect of the effect of one outer boundary on the probabilities of reaching the other; the use of diffusion theory to represent a discrete sampling process; and the application of normal distribution theory to a problem involving a binomial variate. As a check on the validity of the theory, some exact probabilities have been calculated for the procedure tabulated in Table 5, with $\alpha = 0{\cdot}025$, $\beta = 0{\cdot}05$, $\theta_1 = 0{\cdot}8$, $N_1 = 44$. The boundaries, including the modified middle boundary of M', are shown in Fig. 2. The probabilities of reaching the various boundary points, for $\theta = 0{\cdot}5$ and $0{\cdot}8$, are given in Table 6.

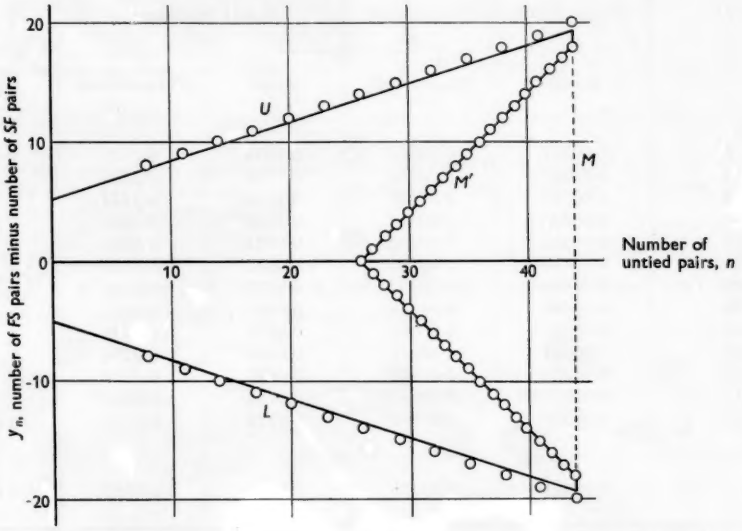


Fig. 2. Comparison of two proportions: boundaries for procedure with $\alpha = 0{\cdot}025$, $\beta = 0{\cdot}05$, $\theta_1 = 0{\cdot}8$ and $N = 44$. Boundary points shown by circles.

The exact probabilities were calculated by enumeration of paths (Stockman & Armitage, 1946), for various values of θ. As in (9), the diffusion approximation gives, for the probability of crossing U before $n = n'$,

$$1 - F\left(\frac{a - mn'}{\sigma\sqrt{n'}}\right) + \exp\left(\frac{2am}{\sigma^2}\right) F\left(\frac{-a - mn'}{\sigma\sqrt{n'}}\right), \tag{10}$$

where $a = 5{\cdot}248$, $\sigma^2 = 4\theta(1-\theta)$, and $m = 2\theta - 1{\cdot}3219$.

Table 6 compares the theoretical and actual probability distributions on the thirteen boundary points of U, for $\theta = 0{\cdot}8$ and $0{\cdot}5$. The cumulative distributions are shown in Fig. 3. In these comparisons the theoretical probability of crossing between two adjacent

boundary points n_1' and n_2' is taken as an approximation to the actual probability for the higher value, n_2'; the cumulative distributions therefore give the theoretical and exact probabilities of crossing U in not more than n' observations, for various n'. The agreement seems to be fairly good for $\theta = 0{\cdot}8$, although there is a noticeable excess of theoretical over exact values for $\theta = 0{\cdot}5$. This is to some extent the effect (familiar in Wald's procedures)

Table 6. *Procedure for comparison of two proportions, with* $\alpha = 0{\cdot}025$, $\beta = 0{\cdot}05$ *and* $\theta_1 = 0{\cdot}80$. *Probability distribution on each boundary, for* $\theta = 0{\cdot}5$ *and* $0{\cdot}8$, *together with the theoretical approximation to the distribution on the upper boundary, U, given by* (10)

Probability of crossing boundary at n'th untied pair

n'	$\theta = 0{\cdot}5$		$\theta = 0{\cdot}8$		
	Upper and lower boundaries		Upper boundary		Lower boundary (exact)
	Exact	Theoretical	Exact	Theoretical	
8	0·00391	*0·00871*	0·1678	*0·1366*	$0{\cdot}0^5 00$
11	0·00391	*0·00565*	0·1718	*0·1515*	00
14	0·00317	*0·00457*	0·1429	*0·1454*	00
17	0·00244	*0·00351*	0·1126	*0·1230*	00
20	0·00185	*0·00266*	0·0873	*0·0987*	00
23	0·00140	*0·00202*	0·0675	*0·0774*	00
26	0·00106	*0·00154*	0·0523	*0·0600*	00
29	0·00080	*0·00118*	0·0407	*0·0464*	00
32	0·00061	*0·00091*	0·0318	*0·0358*	00
35	0·00047	*0·00070*	0·0250	*0·0276*	00
38	0·00036	*0·00054*	0·0197	*0·0213*	01
41	0·00028	*0·00043*	0·0156	*0·0165*	07
44	0·00022	*0·00034*	0·0124	*0·0129*	$0{\cdot}0^5 26$
Total, $n' \leqslant 44$	0·02047	*0·03275*	0·9473	*0·9532*	$0{\cdot}0^5 34$
	Middle boundary (exact)		Middle boundary (exact)		
39–44	0·0090	upper section	0·0270		
33–38	0·0824		0·0197		
27–32	0·3108		0·0053		
26	0·1546		0·0005		
27–32	0·3108	lower section	0·0001		
33–38	0·0824		0·0000		
39–44	0·0090		0·0000		
Total, $26 \leqslant n' \leqslant 44$	0·9590		0·0526		

of the overshooting of the boundaries, which ensures that at each boundary point the likelihood ratio of H_+ to H_0 exceeds the nominal value of $(1-\beta)/\alpha$.*

The total probabilities, theoretical and actual, of reaching U for various values of θ are shown in Table 7. The agreement is again fairly good. Table 7 gives also the exact average

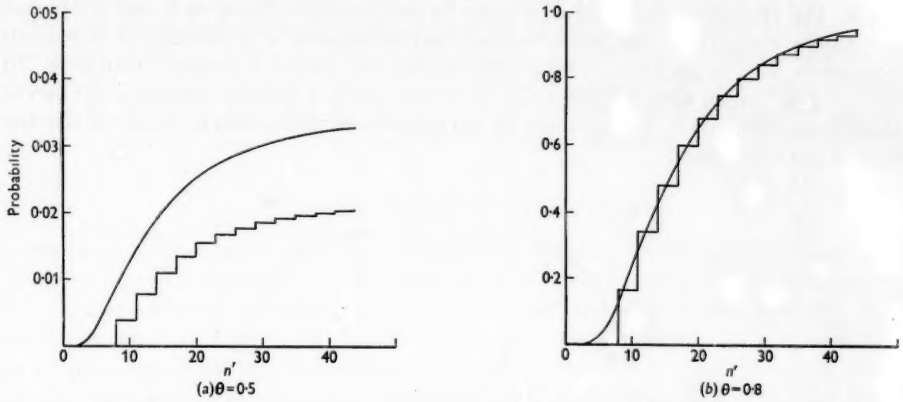


Fig. 3. Comparison of two proportions: procedure with $\alpha = 0{\cdot}025$, $\beta = 0{\cdot}05$, $\theta_1 = 0{\cdot}8$ and $N = 44$. Probability of reaching upper boundary with not more than n' untied pairs, (*a*) when $\theta = 0{\cdot}5$ and (*b*) when $\theta = 0{\cdot}8$. The step function shows exact values, and the continuous function the theoretical diffusion approximation.

Table 7. *Procedure for comparison of two proportions, with* $\alpha = 0{\cdot}025$, $\beta = 0{\cdot}05$, $\theta_1 = 0{\cdot}80$, *and* $N = 44$. *Exact and theoretical probabilities of reaching* U, *and exact average and variance of sample number, for various values of* θ

θ	Probability of reaching U		Exact distribution of sample number (untied pairs)	
	Exact	Theoretical	Mean	Variance
0·50	0·0205	*0·0327*	29·21	19·83
0·60	0·1490	*0·1850*	30·09	46·06
0·70	0·5623	*0·5860*	27·90	115·55
0·80	0·9473	*0·9532*	18·81	97·21
0·85	0·9941	*0·9965*	14·46	49·42
0·90	0·9999	*1·0000*	11·43	18·80
0·95	1·0000	*1·0000*	9·41	5·57

and variance of the sample number for various values of θ; these relate to the procedure with the modified boundary, M'. The ASN curve is shown in Fig. 4, together with the approximate ASN curve for a two-sided modification of a Wald sequential procedure with the same nominal values of α, β and θ_1, and the equivalent fixed-sample-size of $N_0 = 30$.

* A very much better approximation to the probabilities for $\theta = 0{\cdot}5$, is obtained by calculating the likelihood ratio of H_+ to H_0 at each boundary point, and dividing the theoretical probability for $\theta = 0{\cdot}8$ by this ratio.

(In the fixed-sample-size procedure with $N_0 = 30$ the boundaries U, M and L correspond to the results $n_1 \geqslant 21$, $10 \leqslant n_1 \leqslant 20$, and $n_1 \leqslant 9$, respectively; the probabilities of reaching U on H_0 and H_+ are then 0·021 and 0·939, which agree reasonably well with the corresponding values in Table 7.)

3·5. The procedures described above may be used for other situations in which repeated qualitative comparisons are made between two treatments. For example, two medical treatments may be placed in order of effectiveness for each of a number of subjects. An 'untied pair' would then correspond to a subject giving a definite preference for one or other treatment, and n_1 and n_2 would be the numbers of preferences in favour of the first and second treatments.

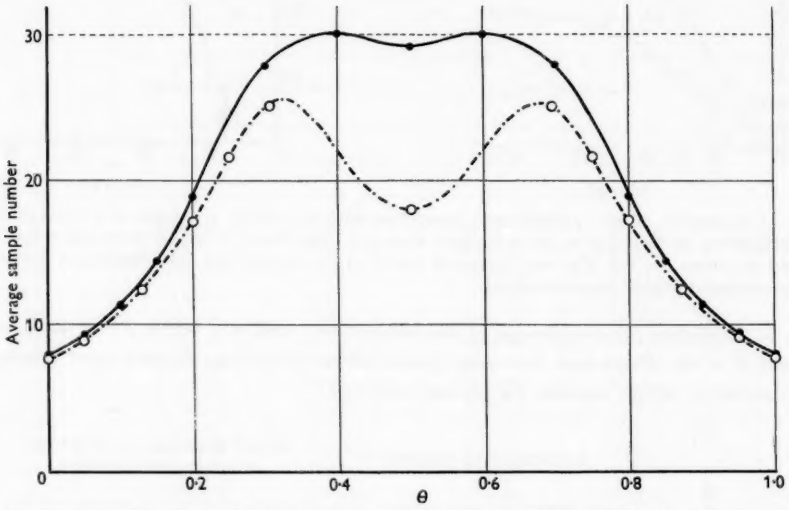


Fig. 4. Comparison of two proportions: average sample number (ASN) of untied pairs as a function of θ, for three procedures with approximately $\alpha = 0{\cdot}025$, $\beta = 0{\cdot}05$, $\theta_1 = 0{\cdot}8$. -·-·- Two-sided Wald scheme (Armitage, 1954; ASN for $\theta = 0{\cdot}5$ from de Boer's (1953) method, other values approximated by formula for one-sided scheme (Wald, 1947)). —— Restricted sequential procedure with $N = 44$ (exact values). ------ Fixed sample-size procedure with $N_0 = 30$.

4. Normal distribution with unknown variance

A restricted procedure based on the sequential t-test could theoretically be obtained by using the upper boundary of a two-sided sequential t-test (National Bureau of Standards, 1951), and terminating the procedure after a fixed number, N, of observations, choosing N in such a way that the probability, on the null hypothesis, of reaching the upper boundary took some specified value. In practice, no method of determining N appears to be available, since nothing is known of the distribution of sample size on sequential t boundaries.

Suppose that observations x_i are drawn from a normal distribution with mean μ and variance σ^2. The sequential procedure is based on the statistic

$$z_n = \left(\sum_{i=1}^{n} x_i\right)^2 \Big/ \left(\sum_{i=1}^{n} x_i^2\right).$$

Suppose that a restricted sequential procedure is required, with an upper boundary U and a terminating boundary M (at $n = N$), satisfying the conditions:

(a) on the hypothesis H_0 that $\mu = 0$, the probability that sampling will end on U is 2α;

(b) on the hypothesis $H_+[H_-]$ that $\mu/\sigma = \Delta > 0\,[\mu/\sigma = -\Delta < 0]$, the probability of reaching U is $1-\beta$.

If U is chosen to be the upper boundary of a sequential t-test, with risks 2α and β and the appropriate value for Δ, the probabilities required by (a) and (b) will be approximately in the correct ratio. In the absence of any sound method of determining N, one might determine the size of the equivalent fixed sample, say N_0, and conjecture that N/N_0 takes approximately the same value as when σ^2 is known. This value is given in terms of α and β in Table 2.

5. SIGNIFICANCE TESTING AND ESTIMATION

In the literature on sequential tests of significance, it has almost invariably been assumed that the object of a significance test is to state whether or not a null hypothesis is disproved at some fixed level of probability (say, 5%). In fixed-sample-size tests, however, it is usually regarded as more useful to indicate the exact probability at which a particular result is just significant; that is, the probability of obtaining, in repeated sampling under specified conditions, a result at least as extreme (in some sense) as that observed. There seems no reason why this practice should not be followed with sequential experimentation, provided that the boundary points can be ordered, in some reasonable way, in terms of the apparent deviation from the null hypothesis.

In the type of sequential procedure considered in this paper, such an ordering can be made with little ambiguity. Suppose that the procedure described in § 2 is used to test the null hypothesis that $\mu = 0$. If there were no overshooting of the boundaries the sample mean, $\bar{x}$, when the procedure terminated, would increase monotonically as the boundary point moved round the boundaries from low to high values of n on L, along M, and from high to low values of n on U. If $|\bar{x}|$ is used as an estimate of departure from the null hypothesis, the probability level corresponding to any boundary point, on a two-sided significance test, will be the probability (on the null hypothesis) of reaching boundary points with higher values of $|\bar{x}|$ than that observed.

Some ambiguity is caused by the discrete steps in n, but it would be easy to formulate reasonable ordering rules. For example, if the procedure terminated at a sample size n', the possible pairs of $(n', \bar{x})$ could be arranged in decreasing order of significance as follows: (i) all results with $n' = 1$, in decreasing order of $|\bar{x}|$; (ii) all results with $n' = 2$, in decreasing order of $|\bar{x}|$; and so on. Alternatively, results could be arranged in order of $|\bar{x}|$ without reference to n'.

In the binomial procedures of § 3, also, slight ambiguity is caused by the overshooting of the boundaries, so that the proportion of FS pairs at the termination, $\hat{\theta}$, may not always increase monotonically as the boundary point moves from one extreme to the other. The simplest rule here would be to arrange the points in terms of the natural order round the boundaries. Thus, on the outer boundaries, smaller values of n' would represent more extreme results, and on the modified middle boundary the results would be ordered according to $|\hat{\theta} - \frac{1}{2}|$. As is shown by Table 6, the diffusion theory may not provide an adequate approximation to the required probability, and it would be preferable to calculate it exactly.

This approach may be applied to other types of sequential procedure, in particular to Wald sequential tests, provided that the probabilities, on the null hypothesis to be tested, of reaching the individual boundary points are known.

If these probability distributions are known it is possible in principle to examine the properties of point estimators of unknown parameters, and to formulate rules for obtaining confidence intervals. In general, intervals obtained from the usual fixed-sample-size formulae will not yield the nominal confidence coefficients in repeated sampling with sequential stopping rules. Some numerical investigations on a number of sequential procedures for binomial variates will be reported in another paper.

I am indebted to Miss I. Allen for computational help.

REFERENCES

ANSCOMBE, F. J. (1954). Fixed-sample-size analysis of sequential observations. *Biometrics*, **10**, 89–100.

ARMITAGE, P. (1954). Sequential tests in prophylactic and therapeutic trials. *Quart. J. Med.* **23**, 255–74.

BAKER, A. G. (1950). Properties of some tests in sequential analysis. *Biometrika*, **37**, 334–46.

BARTLETT, M. S. (1946). The large-sample theory of sequential tests. *Proc. Camb. Phil. Soc.* **42**, 239–44.

BROSS, I. (1952). Sequential medical plans. *Biometrics*, **8**, 188–205.

DE BOER, J. (1953). Sequential test with three possible decisions for testing an unknown probability. *Appl. Sci. Res.* B, **3**, 249–59.

FELLER, W. (1950). *An Introduction to Probability Theory and its Applications*. New York: Wiley.

GORDON, R. D. (1941). Values of Mills' ratio of area to bounding ordinate and of the normal probability integral for large values of the argument. *Ann. Math. Statist.* **12**, 364–6.

NATIONAL BUREAU OF STANDARDS (1951). *Tables to Facilitate Sequential t-Tests*. Applied Mathematics Series 7. Washington: Government Printing Office.

PAGE, E. S. (1957). On problems in which a change in a parameter occurs at an unknown point. *Biometrika*, **44**, 248–52.

RAO, C. R. (1950). Sequential tests of null hypotheses. *Sankhyā*, **10**, 361–70.

SNELL, E. S. & ARMITAGE, P. (1957). A clinical comparison of diamorphine and pholcodeine as cough suppressants by a new method of sequential analysis. *Lancet* (in the Press).

STOCKMAN, C. M. & ARMITAGE, P. (1946). Some properties of closed sequential schemes. *J. R. Statist. Soc. Suppl.* **8**, 104–12.

WALD, A. (1947). *Sequential Analysis*. New York: Wiley.

ON THEORETICAL MODELS FOR COMPETITIVE AND PREDATORY BIOLOGICAL SYSTEMS

By M. S. BARTLETT
University of Manchester

1. Preamble*

Differences of opinion are obviously possible on the degree to which admittedly over-simplified theoretical models can explain some of the complex observational phenomena to be found in nature. Criticisms from biologists of the mathematical work of Lotka (1925), Volterra (1926) and subsequent writers on the growth and interaction of biological populations have, however, sometimes been justified and sometimes unjustified, for in spite of inevitable limitations such work constitutes a permanent contribution to the understanding of how populations may behave. A significant constructive survey was made by Gause (1934) when he attempted to bridge the gap between theoretical models and natural biological phenomena by controlled laboratory experiments in animal ecology. While experiments also have limitations as representations of nature, the role of both theory and experiment in the physical sciences might be recalled by any biologists inclined to be sceptical of the value of either.

The interrelation of these approaches in biology may be illustrated in the field of epidemiology. Here the vicissitudes of infected populations have been studied in the laboratory as well as in the field; but, as I have emphasized elsewhere (Bartlett, 1956, 1957), the properties of theoretical models indicate, among other things, the extent to which population size may sometimes be crucial in the probable sequence of events, and thus indicate to what extent laboratory observations will have any similarity to larger-scale field observations even if the same model applies to both. An essential point is that recent theoretical formulations explicitly recognize the discrete character of populations and the stochastic or random aspect of changes, as distinct from strictly deterministic formulations. The need for this in ecology, which was already perhaps envisaged by Gause (1934, p. 124), is quite apparent in the experiments by Park with the flour-beetle *Tribolium*, in which one of two species together in a container survived not every time, but with a definite probability (e.g. 30 % of times), that could be estimated by replication and changed by changing the environment (see, for example, Neyman, Park & Scott, 1956).

There is now no mathematical difficulty in the *formulation* of stochastic models (see, for example, Bartlett, 1955*a*, 1956), and such formulations have already been made for typical ecological models by Chin Long Chiang (see Kempthorne *et al.* 1954). The greater intractability of even the simplest of these is, however, a serious obstacle to progress, especially in animal ecology, where even in the deterministic formulations of Lotka and Volterra many simplifications, such as neglect of age structure or of other heterogeneity, were made. One aim of the ensuing discussion is to indicate the enhanced value of deterministic formulations of population dynamics when properly interpreted within more comprehensive stochastic

* Some of these introductory comments were also included in a survey paper, 'Some applications of probability and statistical theory in biology', given at the Third Soviet Mathematical Congress held in Moscow in 1956.

models. After some remarks on the logistic model of population growth for a single species, two problems are referred to: (i) the classical Lotka–Volterra predator-prey relation, (ii) competition between two species, with special reference to the competition between the two species of flour-beetle, *Tribolium confusum* and *Tribolium castaneum*, as investigated by Park.

2. The logistic model

The properties of the logistic law of growth, introduced by Verhulst (see, Andrewartha & Birch, 1954) and much discussed in the literature, need not be recapitulated here. In its deterministic form, the rate of increase of a population of size N is assumed to be

$$DN = \alpha N - \beta N^2, \tag{1}$$

where $D \equiv d/dt$ denotes differentiation with respect to the time t. It is well known that this simple equation ignores deterministic oscillations arising from age structure, and in a complete discussion it would be advisable to investigate the amplitude of oscillations in stochastic formulations, due to such a complication, along the lines to be developed for the simpler case above. However, the main point I wish to make is that the rate of increase given in (1) is a net balance of births and deaths, and many different stochastic models compatible with (1) are possible (cf. Kendall, 1949). In the extreme case deaths are negligible compared with births, and the chance of a birth in time δt (independently of previous events) will be assumed to be $(\alpha N - \beta N^2)\,\delta t + o(\delta t)$.* The difference between this stochastic model and (1) has been already studied in some detail (see Feller, 1939; Bailey, 1950). The asymptotic value for the population size is in this case still fixed at α/β. But if we assume that the chance of a birth in time δt is $(\alpha_1 N - \beta_1 N^2)\,\delta t$ and of a death $(\alpha_2 N + \beta_2 N^2)\,\delta t$, then we have four coefficients connected by the relations

$$\alpha_1 - \alpha_2 = \alpha, \quad \beta_1 + \beta_2 = \beta. \tag{2}$$

In a small interval δt, we may write

$$\delta N = (\alpha N - \beta N^2)\,\delta t + \delta Z_1 - \delta Z_2, \tag{3}$$

where, as the first term on the right-hand side of (3) represents the systematic part of the random or stochastic change δN (which can only be 0 or 1 as $\delta t \to 0$), δZ_1 and δZ_2 are independent (modified) Poisson variables with zero means, and variances $(\alpha_1 N - \beta_1 N^2)\,\delta t$, $(\alpha_2 N + \beta_2 N^2)\,\delta t$ respectively. To illustrate the use of equation (3), consider the asymptotic situation when $N \sim \alpha/\beta$; put, more precisely, $N = \alpha(1+u)/\beta$. Then

$$\delta u = -\alpha(1+u)\,u\,\delta t + \beta(\delta Z_1 - \delta Z_2)/\alpha. \tag{4}$$

This non-linear stochastic equation can be approximated for small u by

$$\delta u = -\alpha u\,\delta t + \delta Z, \tag{5}$$

where δZ has variance $\{(\alpha_1+\alpha_2)\beta/\alpha + \beta_2 - \beta_1\}\,\delta t = \gamma\,\delta t$, say. In any 'steady state' the quantity u at time t has the same statistical properties as the quantity $u + \delta u$ at time $t + \delta t$; this gives for the variance σ_u^2 of u, on squaring and averaging the equation for $u + \delta u$ obtained from (5),

$$\sigma_u^2 = \sigma_u^2(1 - 2\alpha\,\delta t) + \gamma\,\delta t,$$

* This is the simplest probability assumption which for large N is equivalent to (1). The last term $o(\delta t)$, denoting a remainder R such that $R/\delta t \to 0$ as $\delta t \to 0$, is inserted for strict rigour, but is for convenience omitted in subsequent formulae.

whence $\sigma_u^2 = \frac{1}{2}\gamma/\alpha$. Strictly, as the value $N = 0$ represents an 'absorbing state' (and the value $N = \alpha_1/\beta_1$ an upper limit which cannot be exceeded), it follows from the theory of finite Markov chains that the ultimate state is $N = 0$, provided this state is accessible from any other admissible value of N (i.e. α_2 and β_2 positive). However, it seems evident from the result above for σ_u^2 that the chance of extinction, once the value of N is in the neighbourhood of the value α/β, may be neglected for any given time interval, provided γ/α is sufficiently small. Under such conditions the population will thus, in contrast with the pure birth process, continue to show fluctuations with this variance. Clearly, before persistent fluctuations observed in real animal populations are considered incompatible with the logistic model, their size and characteristics should be compared with fluctuations predicted from stochastic models like the above. *It is essential for this comparison, in the reporting of laboratory experiments, that individual replications be recorded separately, and also that the total size of the population (rather than its density) be noted.*

In the discussion of the logistic curve given by Andrewartha & Birch (1954), oscillatory fluctuations in the total population are depicted from experiments on *Tribolium*, the weevil *Calandra oryzae* and the cladoceran genus *Daphnia magna*; and the effect of a more complex life history than assumed for the logistic model is, as the authors note, a possible contributory cause of these. The existence of damped oscillations in a deterministic model with an age structure has been noted by Leslie (1948), and the further important point is that in a stochastic model such oscillations can maintain themselves even in the steady state, with an amplitude depending on the population size. In the simple one age-group case above, the fluctuations have no true oscillatory character; in a multi-stage population, oscillatory tendencies arise if the roots determining the behaviour of small fluctuations about the steady state are complex-valued (such investigation of the nature of fluctuations is illustrated below in the two further problems to be discussed). In the case of bisexual reproduction, the further complication of unequal numbers of the two sexes should also strictly be taken into account.

3. THE CLASSICAL PREY-PREDATOR SYSTEM

The simplest deterministic Lotka–Volterra equations for a prey-predator system are

$$\left.\begin{aligned} DH &= (\alpha_1 - \beta_1 P)H, \\ DP &= (-\alpha_2 + \beta_2 H)P, \end{aligned}\right\} \tag{6}$$

where H denotes the population size of the 'prey' (or 'hosts'), and P the size of the 'predators' (or 'parasites'). When I refer to these equations as classical, no implication is intended that they represent precisely any real biological systems. They are important in representing the simplest theoretical model of prey-predator interaction that can be specified, and as a prelude to further discussion their main consequences are recalled. From (6),

$$\frac{dH}{dP} = \frac{(\alpha_1 - \beta_1 P)H}{(-\alpha_2 + \beta_2 H)P},$$

whence by integration

$$f(H, P) \equiv -\alpha_2 \log H + \beta_2 H - \alpha_1 \log P + \beta_1 P = \text{constant}. \tag{7}$$

The curves represented by (7) are closed cycles. The equilibrium point is given by $P_0 = \alpha_1/\beta_1$, $H_0 = \alpha_2/\beta_2$, but this is neutral, in the sense that there is no damping towards this point if

the system is at any other point on the H, P graph. For small cycles around (H_0, P_0), it is easily shown that the path is the ellipse

$$\alpha_2 h^2 + \alpha_1 p^2 = \text{constant}, \tag{8}$$

where $H = H_0(1+h)$, $P = P_0(1+p)$. For larger cycles, the path is restricted by the axes $H = 0$, $P = 0$, but it is obvious that any stochastic formulation will lead to fluctuations which will be especially important near *either* $H = 0$ *or* $P = 0$, for if either H or P becomes zero the oscillatory character of the system disappears. Thus, while modified deterministic formulations, either by way of more complicated assumptions in models like (6), or (by A. J. Nicholson and V. A. Bailey; see, for example, Bach & Smith, 1941) in terms of discrete generations, have led to unstable systems, we see that even (6), when incorporated into a stochastic model, is ultimately unstable, for the drift due to stochastic fluctuations will lead sooner or later to the total extinction of the predator species, either (i) before the other species or (ii) by starvation due to the extinction of the prey first.

To illustrate this point, a stochastic model was set up compatible with (6). There is the same difficulty as with the logistic model that many different birth- and death-rates are consistent with the same net rates of increase of H and P, but (corresponding roughly with the conditions in some actual situations) α_1 may be interpreted for simplicity as a pure birth-rate for H, and α_2 as a pure death-rate for P in the absence of H. The relation of the second term $\beta_2 HP$ in the second equation with the death-rate $\beta_1 P$ for H is not so immediate as in the case of epidemic theory, where susceptibles when in contact with infecteds turn into more infecteds, but will be assumed to be a consequence of an increase $\beta_2 H$ in the birth-rate for the predators. (The precise assumptions made are given later below.)

It should be added at this stage that the appropriateness of these simplified assumptions are likely to be even less realistic in the case of the type of system now under consideration than in, say, epidemic theory, for the basic new assumption of randomness introduced into the deterministic equations will often be somewhat dubious (cf. Andrewartha & Birch, 1954, p. 415). This will be the case when the predator is attracted to the prey, for example, by sight or smell; even then, however, the concept of random encounters may become less artificial in a natural environment, where the predator may be temporarily outside the range of such an attraction. Of course the modifying effect of distance and spatial variability in numbers is a further important consideration (cf. Bartlett, 1956), but this will not be considered here.

The constants chosen for the illustrative stochastic model were

$$\alpha_1 = 1, \quad \beta_1 = 0{\cdot}1, \quad \alpha_2 = 0{\cdot}5, \quad \beta_2 = 0{\cdot}02,$$

leading to $P_0 = 10$, $H_0 = 25$. An artificial realization was started with $H = 25$, $P = 2$, and developed* by standard 'Monte Carlo' technique (cf., for example, Bartlett, 1955, p. 131) until extinction of either H or P had occurred. The series so obtained is graphed in Figs. 1 and 2. At first it showed little stochastic drift, and in spite of low values of H at 'time' 14, and P at time 8, persisted with similar cycles until the rather sudden switch to a smaller

* To correspond to $\beta_2/\beta_1 = \frac{1}{5}$, the predator population was assumed to increase by one for only one in five times that a prey was exterminated. The precise asymptotic transition probabilities assumed were thus: H, P to

$H+1, P$	$H\,\delta t$
$H, P-1$	$\frac{1}{2}P\,\delta t$
$H-1, P$	$0{\cdot}08HP\,\delta t$
$H-1, P+1$	$0{\cdot}02HP\,\delta t$

and less well-defined cycle and then to final extinction of P. Qualitatively, this series shows many of the characteristics that have been observed in small-scale laboratory experiments—variation in amplitude of the cycle, possible extinction first of H (and then P), or of P directly (and subsequent unlimited increase of H).

A quantitative comparison with suitable laboratory data would be useful, but there are at present difficulties in any very precise comparisons. The first are theoretical, and arise

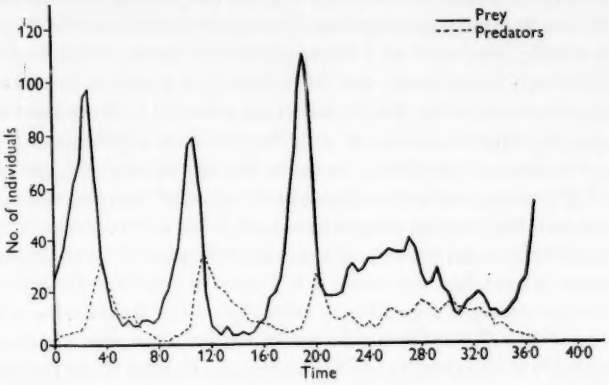


Fig. 1. An artificial realization of a stochastic model for the prey-predator relation.

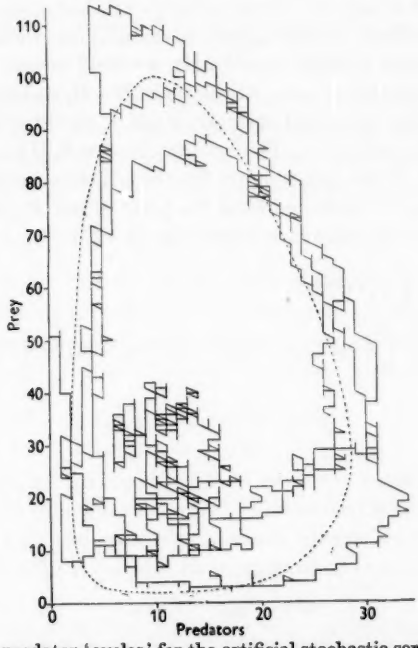


Fig. 2. Prey-predator 'cycles' for the artificial stochastic series. The closed deterministic cycle is also shown (the dotted curve) for comparison.

from the mathematical intractability of a complete solution even of these over-simplified models (it is for this reason that the construction of artificial series is particularly informative). The second arise from the extra complications present in real animal populations. Thus a striking and somewhat unique example of oscillations obtained by Gause (1935), in which the prey were yeast cells *Saccharomyces exiguus* and the predator *Paramecium aurelia*, has been criticized by Andrewartha & Birch (1954, p.440) because, for instance, of the tendency in such experiments for sedimentation of the yeast at the bottom of the culture. While we should bear in mind such objections, we might notice that the data in this experiment appear reasonably consistent with the above simple theory. Thus, from the data given by Gause, we have $\alpha_1 \sim 1$, $\alpha_2 \sim 0{\cdot}45$, and since from the graph it appears that $P_0 \sim 100$, $H_0 \sim 1{\cdot}5 \times 10^7$, we should conclude that $\beta_1 \sim 0{\cdot}01$, $\beta_2 \sim 3 \times 10^{-8}$. From the theory of (small) oscillations about (H_0, P_0), the period of a cycle, which is approximately $2\pi/\sqrt{(\alpha_1 \alpha_2)}$, is calculated to be 9·4 days, a 'prediction' in quite fair agreement with the observed period of about 8 days.* Moreover, while the ultimate extinction of the predator on the stochastic model has been noted, the chances of extinction after one or two cycles may be relatively small (as in the artificial series above), or even microscopically small in certain cases. In spite of the absence of complete solutions, it is known in epidemic theory (Bartlett, 1956) that such chances of extinction depend very critically on the magnitudes of the coefficients occurring in the equations. An approximate device employed in such a context for assessing the order of magnitude of the chances of extinction may be used in the present context also. It consists in neglecting some of the variation in the numbers of the second species when the number of the first species is low and liable to extinction. Thus if in the example under discussion we note that H is large and thus on the theoretical model fairly safe for survival (any real 'sedimentation effect' would support this assumption though a strong dependence between individuals of large colonies would not), we need merely consider the chance of extinction of P at the bottom of its cycle. At this point $H \sim H_0$ and the birth- and death-rates for P are equal. If this situation could maintain itself, P would ultimately become extinct (see, for example, Bartlett, 1955, p. 71). The fact that, for low P, H is increasing will, however, modify such a certainty. If we put roughly for the effective birth-rate $\alpha_2 + c \sin(2\pi t/T)$, where T is the period and t is reckoned from the point of low P, then from known theory (Kendall, 1948) the chance of extinction is given by $(1+1/J)^{-P} \sim e^{-P/J}$, where

$$J = \int_0^t \alpha_2 e^{\rho(x)}\, dx,$$

$$\rho(x) = -\int_0^x c \sin(2\pi y/T)\, dy = (\tfrac{1}{2}cT/\pi)\{\cos(2\pi x/T) - 1\}.$$

This gives, if $b = \frac{1}{2}T/\pi$,

$$J = \alpha_2 \int_0^t e^{bc\{\cos(x/b)-1\}}\, dx.$$

This integral may be evaluated as a series of Bessel functions for general t, but while it is of particular interest to us for t of the order of $\frac{1}{4}T$, it is convenient to remember that the chance of extinction of P must be a strictly increasing function of t, and so we shall not underestimate its value if we choose t to be larger than this, say T. For this value of t

$$J = T\alpha_2 e^{-bc} I_0(bc),$$

* The 'observed' period for the mock series (Fig. 1) suggests that the theoretical formula may still approximately be applied even for quite large oscillations.

where $I_0(x)$ is the Bessel function of zero order (of the first kind, and of imaginary argument). By inspection of the graph given by Gause, c is about $\frac{5}{6}\alpha_2$. Hence, with $\alpha_2 = 0{\cdot}45$ and $T = 9{\cdot}4$, it is found that $J = 2{\cdot}6$, and thus when $P \sim 15$, the chance of extinction during the critical part of the cycle is of order $e^{-5{\cdot}7} \sim 0{\cdot}003$.* Thus while it has been suggested by Andrewartha & Birch (1954, p. 440) that *Paramecium* avoid extinction by growing smaller when no food is available, these results seem at least reasonably compatible with the simple mathematical model.

The relevance of these extinction probabilities in ecology would be more convincing if further data were available where extinction did in fact occur, and calculation gave a high chance of extinction in such cases. First of all it may be checked that this chance is high for the artificial series of Fig. 1. Using the deterministic solution to calculate the order of magnitude of low H or low P after one cycle, with the given initial conditions, the chance of extinction per cycle was calculated to be of the order 0·6 for P and 0·3 for H. However, a corresponding calculation for the experimental data on mites reported by Gause, Smaragdova & Witt (1936) gave a very small chance of extinction per cycle, in contrast with the invariable extinction actually observed. The specific calculation was made for the following series (in wheat flour):

Days	0	6	12	23	27	32	35	38	41	44	47
Prey (*Aleuroglyphus agilis*)	50	24	28	256	408	496	288	32	—	20	—
Predators (*Cheyletus eruditus*)	5	4	12	12	24	64	96	120	44	24	8

The numbers are estimated total numbers at all stages of growth, excluding eggs; quoted from Gause *et al.* (1936).

Only very crude estimates ($\alpha_1 \sim 0{\cdot}08$, $\alpha_2 \sim 0{\cdot}3$, $\beta_1 \sim 0{\cdot}004$, $\beta_2 \sim 0{\cdot}003$) of the coefficients were made from these data, as there appears no doubt from these and similar data that the model is inappropriate because the true cycle is deterministically unstable. The effect of a time lag in the growth of new individuals is one possible explanation of this, and is examined further in the next section.

4. THE EFFECT OF A LAG IN BIRTHS

It is assumed that there is a lag τ_1 in the growth to maturity of the prey, so that (in terms of adults) the increases may approximately for small τ_1 be regarded as due to the prey present at time τ_1 previously. A lag τ_2 is similarly assumed for new predators. This modifies equations (6) to

$$\left.\begin{aligned} DH(t) &= \alpha_1 H(t-\tau_1) - \beta_1 P(t)\, H(t), \\ DP(t) &= -\alpha_2 P(t) + \beta_2 H(t-\tau_2)\, P(t-\tau_2). \end{aligned}\right\} \tag{9}$$

In terms of the differential operator D, equations (9) can be written

$$\left.\begin{aligned} Dh &= \alpha_1 e^{-\tau_1 D}(1+h) - \alpha_1(1+h)\,(1+p), \\ D_p &= -\alpha_2(1+p) + \alpha_2 e^{-\tau_2 D}(1+h)\,(1+p), \end{aligned}\right\} \tag{10}$$

* In this calculation the value 15 for P was taken as typical from the graph; if alternatively, as in the further examples, it is calculated theoretically from the initial conditions, the value obtained for the chance of extinction is even smaller.

where $H = H_0(1+h)$, $P = P_0(1+p)$. For h, p (and τ_1, τ_2) small, equations (10) become approximately

$$\left.\begin{aligned} Dh &\sim -\tau_1\alpha_1 Dh - \alpha_1 p, \\ Dp &\sim \alpha_2 h - \alpha_2\tau_2 D(h+p). \end{aligned}\right\} \tag{11}$$

The character of small oscillations about H_0, P_0 is determined by the equation in D,

$$D^2(1+\tau_1\alpha_1)(1+\tau_2\alpha_2) - \alpha_1\alpha_2(\tau_2 D - 1) = 0,$$

giving rise to instability (for $\tau_2 > 0$) even on the deterministic formulation. The approximate solution for h, p is found to be

$$h \sim (A\cos\theta t + B\sin\theta t)\, e^{\frac{1}{2}\alpha_1\alpha_2\tau_2 t},$$

$$\alpha_1 p \sim (A\theta\sin\theta t - B\theta\cos\theta t)(1+\tau_1\alpha_1)\, e^{\frac{1}{2}\alpha_1\alpha_2\tau_2 t} - \tfrac{1}{2}\alpha_1\alpha_2\tau_2 h,$$

where $\theta^2 \sim \alpha_1\alpha_2(1-\tau_1\alpha_1)(1-\tau_2\alpha_2)$. Thus

$$\alpha_1 p^2(1-\tau_1\alpha_1+\tau_2\alpha_2) + \alpha_1\alpha_2\tau_2 hp + \alpha_2 h^2 \sim C\, e^{\alpha_1\alpha_2\tau_2 t}, \tag{12}$$

where $C \equiv \alpha_2(A^2+B^2)$ is a constant.

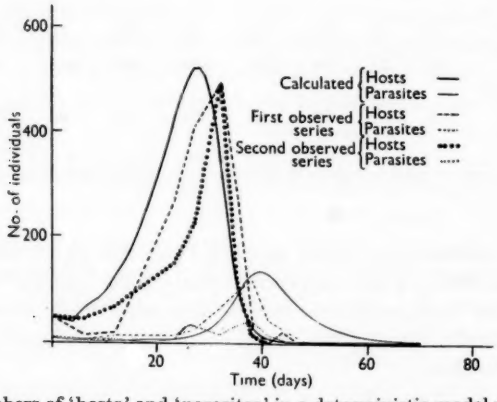


Fig. 3. Numbers of 'hosts' and 'parasites' in a deterministic model with age-lag. Two observed series taken from Gause *et al.* (1936) are shown for comparison.

The chance of extinction would obviously be high if the deterministic number H at the bottom of the cycle is as low as, say, $\frac{1}{2}$. A calculation based on equation (12) (while rather rough, because this equation does not exactly apply except for small h, p) indicated that τ_2 need not be more than about 5 days, suggesting that the above simple modification might be sufficient to bring the assumed model more in line with the data under discussion.

A numerical check on this suggestion was considered advisable, especially as the new model implies a modified growth rate for the hosts (in the absence of the predators) of $\alpha_1' \sim \alpha_1/(1+\tau_1\alpha_1)$. A completely new set of values of the coefficients was estimated, still rather roughly, as $\tau_1 = \tau_2 = 4$, $\alpha_1 \sim 0{\cdot}25$, $\alpha_2 \sim 0{\cdot}15$, $\beta_1 \sim 0{\cdot}015$, $\beta_2 \sim 0{\cdot}002$. The deterministic series calculated for these values are compared in Fig. 3 with the corresponding observed values quoted above. Some apparent discrepancies remain, but it should be noticed that a second observed series with the same initial conditions as the first (also shown in Fig. 3) shows considerable divergence from it. Thus any more detailed analysis, while no doubt

requiring a rather more elaborate study of age structure, would have to allow for such stochastic fluctuations. The immediately relevant point, however, is that an effectively zero value of H at the end of the cycle follows from this slight modification of the model.

5. THE EFFECT OF IMMIGRATION

The stochastic instability of even the classical prey-predator model (6) naturally suggests that the theoretical effect of immigration should be considered (cf. Bartlett, 1956). In simple epidemic models of the mixed susceptible and infected population (with infection by contagion) it is well known that epidemics cannot recur unless the susceptibles are renewed. A more difficult point is to what extent renewal of infecteds from outside is necessary, but recurrence is in any case made certain by such renewals. In the prey-predator system, the immigration of *either* prey *or* predator will prevent absorption at the corresponding axis $H = 0$ or $P = 0$, but absorption at the other is still possible. If, however, damping terms are introduced into the deterministic model by such immigration, extinction may be greatly delayed. Finally, if immigration of *both* prey and predators is allowed, it is obvious that extinction is completely prevented.

It was consequently interesting to find that Gause (1934) had introduced immigration experimentally into his laboratory cultures, to avoid the extinction difficulty. His theoretical reasons seemed, however, rather *ad hoc* and incomplete, for we have seen that with small laboratory populations extinction would not, even in the absence of further complications, be necessarily unexpected on an explicit stochastic formulation.

With immigration, equations (6) are modified to

$$\left.\begin{aligned} DH &= (\alpha_1 - \beta_1 P) H + \epsilon_1, \\ DP &= (-\alpha_2 + \beta_2 H) P + \epsilon_2. \end{aligned}\right\} \tag{13}$$

The deterministic equilibrium point is now given by (H_0, P_0), where

$$\beta_1 \alpha_2 P_0^2 - P_0(\beta_1 \epsilon_2 + \beta_2 \epsilon_1 + \alpha_1 \alpha_2) + \alpha_1 \epsilon_2 = 0. \tag{14}$$

There is one relevant positive root, as

$$(\beta_1 \epsilon_2 + \beta_2 \epsilon_1 + \alpha_1 \alpha_2)^2 > 4 \alpha_1 \alpha_2 \beta_1 \epsilon_2;$$

for small ϵ_1, ϵ_2,

$$P_0 \sim \frac{\alpha_1}{\beta_1} + \frac{\epsilon_1 \beta_2}{\beta_1 \alpha_2}, \quad H_0 \sim \frac{\alpha_2}{\beta_2} - \frac{\epsilon_2 \beta_1}{\beta_2 \alpha_1}.$$

For small oscillations h, p, where $H = H_0(1+h)$, $P = P_0(1+p)$,

$$\left.\begin{aligned} Dh &\sim -h \epsilon_1 \beta_2 / \alpha_2 - \beta_1 P_0 p, \\ Dp &\sim -p \epsilon_2 \beta_1 / \alpha_1 + \beta_2 H_0 h, \end{aligned}\right\} \tag{15}$$

and the equation in D,

$$D^2 + D(\epsilon_1 \beta_2 / \alpha_2 + \epsilon_2 \beta_1 / \alpha_1) + \beta_1 \beta_2 H_0 P_0 \sim 0,$$

implies damping provided ϵ_1 and/or $\epsilon_2 > 0$.

The deterministic damping operates at all amplitudes, at least for small ϵ, ϵ'. For if the function (cf. equation (7))

$$F(H, P) = \beta_1(P - P_0 \log P) + \beta_2(H - H_0 \log H),$$

is differentiated, it is found that

$$DF = -\left\{\frac{\epsilon_1\beta_2}{\alpha_2 H}(\beta_2 H-\alpha_2)^2+\frac{\epsilon_2\beta_1}{\alpha_1 P}(\beta_1 P-\alpha_1)^2\right\}+O(\epsilon_1^3,\epsilon_2^3),$$

so that F always decreases in time until it reaches its minimum value in the neighbourhood of $(\alpha_2/\beta_2, \alpha_1/\beta_1)$, at the point (H_0, P_0).

In view of the instability found in § 4 due to the lag from birth to reproductive age, it is advisable to combine the two effects of such lags and immigration. Since the effects are additive to the first order (for oscillations of small amplitude), it is readily seen that deterministic damping will still be maintained provided

$$\epsilon_1 H_0+\epsilon_2 P_0>\alpha_1\alpha_2\tau_2. \tag{16}$$

As deterministic damping for small ϵ_1, ϵ_2 may be only slight, the amplitude of oscillations may evidently be considerable even when a stochastic 'steady state' (or true stationarity) is feasible.

6. Competition between two species

In the generalization of equation (1) to two species, the simplest model is represented by the two equations

$$\left.\begin{aligned} DN_1 &= (\alpha_1-\beta_{11}N_1-\beta_{12}N_2)N_1,\\ DN_2 &= (\alpha_2-\beta_{21}N_1-\beta_{22}N_2)N_2,\end{aligned}\right\} \tag{17}$$

where N_1 and N_2 refer to the two species. A further simplification is obtained if it is assumed that the restrictive action of the populations is sufficiently well represented by a combined population size $N\equiv N_1+\lambda N_2$, so that $\beta_{12}/\beta_{11}=\beta_{22}/\beta_{21}=\lambda$ (see, for example, Kostitzin, 1939, p. 124). Equations (17) then provide equations*

$$\left.\begin{aligned} DN_1 &= (\alpha_1-\beta_1 N)N_1,\\ DN_2 &= (\alpha_2-\beta_2 N)N_2,\end{aligned}\right\}$$

where $\beta_1=\beta_{11}$, $\lambda\beta_2=\beta_{22}$, whence

$$\left.\begin{aligned} D(\beta_2\log N_1-\beta_1\log N_2) &= \alpha_1\beta_2-\alpha_2\beta_1,\\ N_1^{\beta_2}/N_2^{\beta_1} &= (n_1^{\beta_2}/n_2^{\beta_1})\,e^{(\alpha_1\beta_2-\alpha_2\beta_1)t},\end{aligned}\right\} \tag{18}$$

where n_1 and n_2 are the initial values of N_1 and N_2. This result implies that N_1 or N_2 tends to zero according as $\alpha_1\beta_2$ is less than or greater than $\alpha_2\beta_1$, and the remaining species obeys the single logistic equation (1). This deterministic result is unlikely to be affected by stochastic fluctuations unless $\alpha_1\beta_2\sim\alpha_2\beta_1$, but in this case they may sometimes be important. In particular, suppose the two species refer strictly to two different genealogical lines of the *same* species, so that all the corresponding coefficients in equations (1) are equal. Then $D(N_1-N_2)=0$ when $N_1=N_2$ initially, and with no deterministic damping in N_1-N_2, stochastic drift will take place until *either* N_1 *or* N_2 is zero. This point is noted because it has some relevance to the extinction phenomenon found experimentally in Park's experiments with *Tribolium*, although a more appropriate model will be assumed for this case.

If $\beta_{12}/\beta_{11}\neq\beta_{22}/\beta_{21}$, the equations (17) give a possible equilibrium point (with $N_1, N_2>0$)

$$N_1=\frac{\alpha_1\beta_{22}-\alpha_2\beta_{12}}{\beta_{11}\beta_{22}-\beta_{12}\beta_{21}},\quad N_2=\frac{\alpha_2\beta_{11}-\alpha_1\beta_{21}}{\beta_{11}\beta_{22}-\beta_{12}\beta_{21}},$$

* It should be noted that (in contradiction of a claim by Andrewartha & Birch (1954, p. 411)), there is no logical inconsistency in DN_1 or DN_2 being negative. The terms on the right of the equations are a balance of births and deaths, and will give a negative total if the deaths predominate.

which is found to be stable if

$$\alpha_1\beta_{22} > \alpha_2\beta_{12}, \quad \alpha_1\beta_{21} < \alpha_2\beta_{11},$$

and unstable if

$$\alpha_1\beta_{22} < \alpha_2\beta_{12}, \quad \alpha_1\beta_{21} > \alpha_2\beta_{11}.$$

The last case is interesting, as it implies on the deterministic model that which species survives depends on the initial conditions; hence in a complete stochastic model random variation at the beginning of population growth might be a crucial factor in the survival of one or other species.*

The *Tribolium* model assumed follows the lines suggested by Neyman *et al.* (1956), where the competition between the two species, *T. castaneum* and *T. confusum*, is interpreted more as a predator-prey interaction than a direct competition for food. The four stages of growth are for simplicity condensed into two, referred to as 'passive' and 'active' (the adult stage), and the control of numbers arises apparently from a cannibalistic propensity for the adult to consume the young or 'passive'. This cannibalism appears to apply equally to the devouring of the young of either species, though the adult propensities may differ for the two species.

The Neyman, Park and Scott model was formulated in terms of discrete generations. The general type of stochastic formulation adopted in the present paper can include such discrete generation models (cf. § 4), but the essential points seem to emerge from the simplest stochastic formulation in continuous time of the type represented by equation (3), and this will be used here. The single species case is of some interest for comparison with the straightforward logistic model, and is considered first. Spatial 'diffusion' of the beetle is ignored, though it may be an important further factor in the actual experiments.

The deterministic form of the model assumed is

$$\left.\begin{aligned} DP &= -\mu AP - \nu P + \lambda A, \\ DA &= \nu P - \epsilon A, \end{aligned}\right\} \tag{19}$$

where P is the number of passive, A of active, beetles, and λ and ϵ ($< \lambda$) are birth- and death-rates, ν is the transition rate from passive to active, and μ is a 'voracity' coefficient. The equilibrium point is

$$P_0 = (\lambda - \epsilon)/\mu, \quad A_0 = \nu(\lambda - \epsilon)/(\mu\epsilon). \tag{20}$$

For small oscillations, put $P = P_0(1+p)$, $A = A_0(1+a)$. Then

$$\begin{aligned} Dp &= -\lambda\nu p/\epsilon + \nu a - \nu(\lambda - \epsilon)\, ap/\epsilon \\ &\sim -\lambda\nu p/\epsilon + \nu a, \\ Da &= \epsilon(p - a); \end{aligned}$$

and the equation in D,

$$D^2 + D(\epsilon + \lambda\nu/\epsilon) + (\lambda - \epsilon)\,\nu = 0,$$

indicates damped oscillations. Note that for the function

$$f(p, a) = p - \log(1+p) + \nu[a - \log(1+a)]/\epsilon,$$

$$\begin{aligned} Df &= -\frac{\lambda\nu p^2(1+a)}{\epsilon(1+p)} + \frac{\nu a(ap + 2p - a)}{1+a} \\ &\leqslant -\frac{\nu p^2(1+a)}{1+p} + \frac{\nu a(ap + 2p - a)}{1+a} = -\frac{\nu(a-p)^2}{(1+p)(1+a)}, \end{aligned}$$

* The possibility of such an effect occurring with *Tribolium* was raised by Dr P. H. Leslie in the discussion following a seminar I gave at Oxford on 15 May 1956.

so that f decreases in time, implying that deterministic damping operates at large as well as small amplitudes.

The exact stochastic equations assumed in place of (19) are (cf. equation (3))

$$\left.\begin{aligned}\delta P &= (-\nu P+\lambda A-\mu AP)\,\delta t-\delta Z_1+\delta Z_2-\delta Z_3,\\ \delta A &= (\nu P-\epsilon A)\,\delta t+\delta Z_1-\delta Z_4,\end{aligned}\right\} \tag{21}$$

where the variances of δZ_1, δZ_2, δZ_3 and δZ_4 are indicated by the respective deterministic terms $\nu P\,\delta t$, $\lambda A\,\delta t$, etc. With $P = P_0(1+p)$, $A = A_0(1+a)$ as before, and in the case of small fluctuations,

$$\left.\begin{aligned}\delta p &\sim (-\lambda\nu p/\epsilon+\nu a)\,\delta t+(\delta Z_2-\delta Z_1-\delta Z_3)/P_0,\\ \delta a &\sim \epsilon(p-a)\,\delta t+(\delta Z_1-\delta Z_4)/A_0.\end{aligned}\right\} \tag{22}$$

While presumably on the above model extinction of the population will occur after a long enough time,* this may (for a deterministic 'ceiling' population not too small, but fluctuations relatively small) be so long delayed as to be negligible and an effective or quasi-stationarity be established. Under such conditions, equations (22) yield, on squaring or multiplying together the expressions for $p+\delta p$, $a+\delta a$, and averaging,

$$\begin{aligned}0 &\sim -2\lambda\nu\sigma_p^2/\epsilon+2\nu\operatorname{cov}(p,a)+(\lambda A_0+\nu P_0+\mu A_0P_0)/P_0^2,\\ 0 &\sim 2\epsilon\operatorname{cov}(p,a)-2\epsilon\sigma_a^2+(\nu P_0+\epsilon A_0)/A_0^2,\\ 0 &\sim -\frac{\lambda\nu}{\epsilon}\operatorname{cov}(p,a)+\nu\sigma_a^2+\epsilon\sigma_p^2-\epsilon\operatorname{cov}(p,a)-\nu P_0/(P_0A_0),\end{aligned}$$

whence

$$\left.\begin{aligned}\sigma_p^2 &\sim \frac{\lambda(\lambda\nu+\epsilon^2+\epsilon\nu)}{P_0(\lambda-\epsilon)(\lambda\nu+\epsilon^2)},\quad \sigma_a^2 \sim \frac{\epsilon^2(\lambda-\epsilon)+\lambda^2\nu)}{A_0(\lambda-\epsilon)(\lambda\nu+\epsilon^2)},\\ \operatorname{cov}(p,a) &\sim \frac{\lambda\epsilon^2}{P_0(\lambda-\epsilon)(\lambda\nu+\epsilon^2)}.\end{aligned}\right\} \tag{23}$$

When the two *Tribolium* species are put together, they are denoted by suffices 1 and 2, and the following (deterministic) equations assumed.

$$\left.\begin{aligned}DP_1 &= -\mu_1P_1A_1-\mu_2P_1A_2-\nu_1P_1+\lambda_1A_1,\\ DA_1 &= \nu_1P_1-\epsilon_1A_1,\\ DP_2 &= -\mu_2P_2A_2-\mu_1P_2A_1-\nu_2P_2+\lambda_2A_2,\\ DA_2 &= \nu_2P_2-\epsilon_2A_2.\end{aligned}\right\} \tag{24}$$

An attempt to find a single equilibrium point yields the two (in general) incompatible equations

$$\left.\begin{aligned}-\frac{\mu_1\nu_1}{\epsilon_1}P_1-\frac{\mu_2\nu_2}{\epsilon_2}P_2-\nu_1+\frac{\lambda_1\nu_1}{\epsilon_1} &= 0,\\ -\frac{\mu_1\nu_1}{\epsilon_1}P_1-\frac{\mu_2\nu_2}{\epsilon_2}P_2-\nu_2+\frac{\lambda_2\nu_2}{\epsilon_2} &= 0,\end{aligned}\right\} \tag{25}$$

these representing *parallel* lines in the P_1, P_2 plane (cf. the discussion in Neyman *et al.* 1956).

* There is strictly no *stochastic* upper limit to the population in contrast with the simpler model assumed in § 2, though the effective limit implied by the non-linear 'voracity' term suggests that the ultimate extinction property should still hold.

Before attempting to consider the problem of stability in relation to the zone thus defined, it is instructive to consider the case when the two species become identical in behaviour. In this case, if

$$\left.\begin{aligned} P_1+P_2 &= X, \quad A_1+A_2 = Y, \\ P_1-P_2 &= U, \quad A_1-A_2 = V, \end{aligned}\right\}$$

$$\left.\begin{aligned} DX &= -\mu XY-\nu X+\lambda Y, \\ DY &= \nu X-\epsilon Y, \end{aligned}\right\} \tag{26}$$

where $\lambda_1 = \lambda_2 = \lambda$, etc., and

$$\left.\begin{aligned} DU &= -\mu UY-\nu U+\lambda V, \\ DV &= \nu U-\epsilon V. \end{aligned}\right\} \tag{27}$$

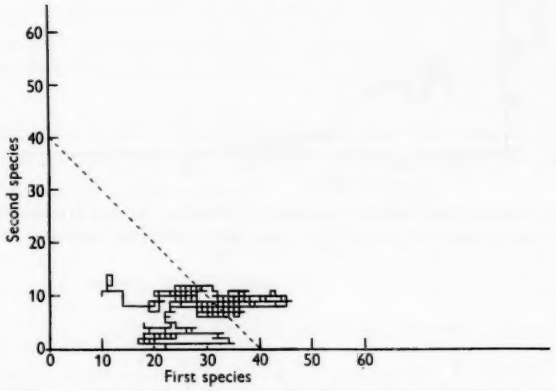


Fig. 4. Graph of total numbers of two competing *Tribolium* species in a simplified stochastic model—symmetric case (the dotted line is the 'equilibrium' line).

The first pair of equations in X and Y are (as they must be) identical with (19), and refer merely to the total population. As Y is damped, it may approximately, even in the stochastic model if numbers are not too small, be replaced by its equilibrium value $\nu(\lambda-\epsilon)/(\mu\epsilon)$. Then equations (27) become

$$\begin{aligned} DU &= -\frac{\lambda\nu}{\epsilon}U+\lambda V, \\ DV &= \nu U-\epsilon V, \end{aligned}$$

with *any* equilibrium solution $U_0/V_0 = \epsilon/\nu$. Hence in the stochastic model there will be random 'drift' along the line $U = \epsilon V/\nu$ until extinction of one or other of the two species occurs. This is not altogether unexpected, for these symbols now refer to individual genealogical lines of the *same* species. The case of two similar species should not be too different from this limiting symmetrical case, for which the chance of extinction of either species first, for symmetrical initial conditions, must be $\frac{1}{2}$. However, if in equations (24) the equilibrium ratios $A_1/P_1 = \nu_1/\epsilon_1$, $A_2/P_2 = \nu_2/\epsilon_2$ are inserted, the incompatible equilibrium equations (25) are replaced by

$$\left.\begin{aligned} D\log P_1 &= -\beta_1 P_1-\beta_2 P_2+\alpha_1, \\ D\log P_2 &= -\beta_1 P_1-\beta_2 P_2+\alpha_2, \end{aligned}\right\} \tag{28}$$

where $\alpha_1 = \lambda_1 \nu_1/\epsilon_1 - \nu_1$, $\beta_1 = \mu_1 \nu_1/\epsilon_1$, etc. a particular case of the competition equations (17). From (28),

$$D \log (P_1/P_2) = \alpha_1 - \alpha_2,$$

and P_1/P_2 tends to increase or decrease according as α_1 is greater or less than α_2. Hence there will be a tendency, as the differences in the characteristics of the two species increase,

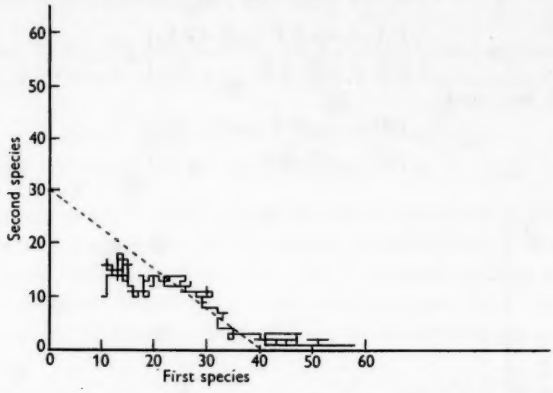


Fig. 5. Graph of total numbers of two competing *Tribolium* species in a simplified stochastic model—asymmetric case (the dotted line joins the equilibrium points on the two axes).

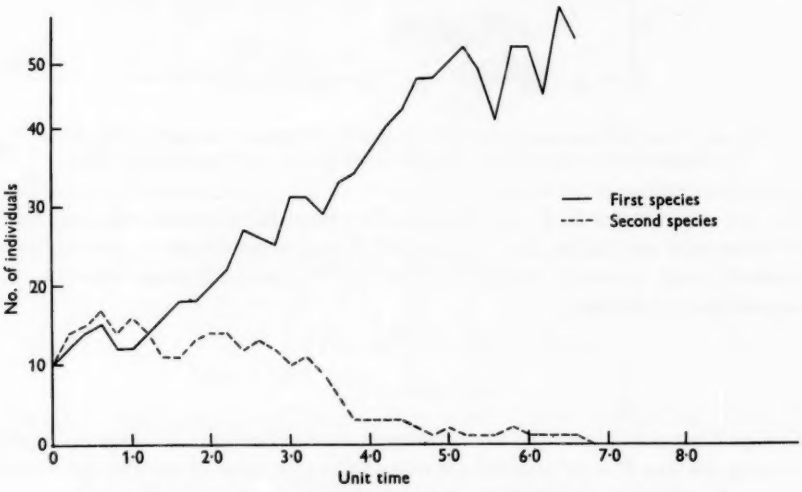


Fig. 6. Total numbers of individuals of the two competing species plotted against time (arbitrary units) in the asymmetric stochastic model.

for the superiority of one species to become more marked and lead more frequently to the survival of that species.

A preliminary illustration of these last conclusions is depicted in Figs. 4 and 5. In the first a Monte Carlo realization is shown (for the total numbers of each species) for the symmetrical case, with the rather arbitrary values $\lambda = 2$, $\mu = 0{\cdot}05$, $\nu = 1$, $\epsilon = 1$, beginning with

$A_1 = A_2 = 10$. The 'equilibrium' line along which stochastic drift occurs is shown dotted. (The corresponding time realization has not been completed, but an idea of the relative time elapsing before extinction of one species occurred is obtained from the number of steps represented in the graphs, viz. 1224 for Fig. 4 and 652 for Fig. 5.) In the case shown in Fig. 5, λ_2 was altered to 3, and ϵ_2 to 2 (with λ_1 still 2, $\epsilon_1 = 1$), and the bias towards extinction of species 1 thus created ($\alpha_1 > \alpha_2$) was supported by the corresponding (and more rapid) extinction in the realization. In this case the time realization is also shown, in Fig. 6.

No quantitative comparison with actual *Tribolium* data has been envisaged at this stage; this will require further investigation. The main object in this preliminary discussion has been to demonstrate qualitatively some of the observational features. Even with the present simple *Tribolium* model (adapted from that formulated by Neyman, Park and Scott) the times to extinction need considering further. It has been suggested above that the duration of any experiment is short compared with the extinction time of a single species, in contrast with the extinction time of one or other of the two competing species, but obviously more definite orders of magnitude for these extinction times would be useful.

I am very grateful to David G. Kendall for some critical comments on my first draft of this paper. I am also indebted to Mrs L. Linnert and Miss C. Caley for assistance with the construction of the artificial realizations shown in the figures.

REFERENCES*

ANDREWARTHA, H. G. & BIRCH, L. C. (1954). *The Distribution and Abundance of Animals.* Chicago.

BACH, P. DE & SMITH, H. S. (1941). Are population oscillations inherent in the host-parasite relation? *Ecology*, **22**, 363–9.

BAILEY, N. T. J. (1950). A simple stochastic epidemic. *Biometrika*, **37**, 193–202.

BARTLETT, M. S. (1955). *An Introduction to Stochastic Processes.* Cambridge University Press.

BARTLETT, M. S. (1956). Deterministic and stochastic models for recurrent epidemics. *Proc. Third Berkeley Symposium on Mathematical Statistics and Probability*, **4**, 81–109. University of California Press.

BARTLETT, M. S. (1957). Measles periodicity and community size. *J. R. Statist. Soc.* A, **120** (in the Press).

ELTON, C. S. (1955). Natural control of animal populations. Review of Andrewartha & Birch (1954), *Nature, Lond.*, **176**, 619.

FELLER, W. (1939). Die Grundlagen der Volterraschen Theorie des Kampfes ums Dasein in wahrscheinlichkeitstheoretischer Behandlung. *Acta biotheor., Leiden*, **5**, 11–40.

FELLER, W. (1940). On the logistic law of growth and its empirical verifications in biology. *Acta biotheor., Leiden*, **5**, 51–66.

GAUSE, G. F. (1934). *The Struggle for Existence.* Baltimore.

GAUSE, G. F. (1935). Experimental demonstration of Volterra's periodic oscillations in the numbers of animals. *J. Exp. Biol.* **12**, 44–8.

GAUSE, G. F., SMARAGDOVA, N. P. & WITT, A. A. (1936). Further studies of interaction between predators and prey. *J. Anim. Ecol.* **5**, 1–18.

KEMPTHORNE, O. *et al.* (Editors) (1954). *Statistics and Mathematics in Biology.* Ames, Iowa: Iowa State College Press.

KENDALL, D. G. (1948). On the generalized 'birth-and-death' process. *Ann. Math. Statist.* **19**, 1–15.

KENDALL, D. G. (1949). Stochastic processes and population growth. *J. R. Statist. Soc.* B, **11**, 230–64.

LESLIE, P. H. (1948). Some further notes on the use of matrices in population mathematics *Biometrika*, **35**, 213–45.

* Including some papers not mentioned in the text.

LESLIE, P. H. (1957). An analysis of the data for some experiments carried out by Gause with populations of the Protozoa *Paramecium aurelia* and *Paramecium caudata*. *Biometrika*, **44**, (to appear).

LOTKA, A. J. (1925). *Elements of Physical Biology*. Baltimore.

MCKENDRICK, A. G. (1926). Applications of mathematics to medical problems. *Proc. Edinb. Math. Soc.* **44**, 98–130.

NEYMAN, J., PARK, T. & SCOTT, E. L. (1956). Struggle for existence. The Tribolium model: biological and statistical aspects. *Proc. Third Berkeley Symposium on Mathematical Statistics and Probability*, **4**, 41–79. University of California Press.

PRATT, D. M. (1943). Analysis of population development in *Daphnia* at different temperatures. *Biol. Bull., Woods Hole*, **85**, 116–40.

SKELLAM, J. G. (1955). The mathematical approach to population dynamics. *The Numbers of Man and Animals* (ed. by Cragg, J. B. & Pirie, N. W.) 31–46. Edinburgh: Oliver and Boyd.

VOLTERRA, V. (1926). Variazioni e fluttuazioni del numero d'individui in specie animali conviventi. *Mem. Acad. Lincei Roma*, **2**, 31–113.

THE CONSISTENCY AND ADEQUACY OF THE POISSON–MARKOFF MODEL FOR DENSITY FLUCTUATIONS

By V. T. PATIL

University of Manchester

SUMMARY. The Poisson–Markoff or emigration-immigration process (as defined in §1·2) has been used as an approximate model for the number of independently moving particles present in particular regions of space, for example, as an approximate model in spermatozoa studies. The postulation of such a model, valid for all modes of division of the space, is shown to be inconsistent in the sense of leading to contradictions, except when τ, the total interval of time considered, is small; the assumptions (cf. §1·3) about the infinitesimal transition probabilities then ensure the consistency of the model to $O(\tau)$. To study the adequacy of the Poisson–Markoff model for some data of spermatozoa counts, the asymptotic χ^2-goodness of fit test based on serial correlations is employed. The goodness of fit tests indicate a significant departure of the spermatozoa data from the Poisson–Markoff model, the departure being not so striking in the uniregional case as in the multiregional case.

Part I

1·1. Consider a system of particles performing some type of stochastic motion in an 'infinite' space. Let $R_1, R_2, \ldots, R_k$ (any k finite non-overlapping regions in the space) and the complementary set R^* represent a mode of division of the space, and let $N_{t,r}$ $(t>0)$ be the number of particles in the region R_r at time t, $\mathbf{N}_t$ being the column vector with $N_{t,r}$ $(r=1,2,\ldots,k)$ as its elements. Then the full specification of the stochastic motion is sufficient to specify the process $\{\mathbf{N}_t\}$. However, not only is the converse not true, but also there may not always be a type of stochastic motion leading to the specified process $\{\mathbf{N}_t\}$. In §1·5, it will be shown that the postulation of the Poisson–Markoff or emigration-immigration process for $\mathbf{N}_t$ (cf. §1·2), when assumed to be valid for all modes of division of the space (as has been used in connexion with spermatozoa studies) leads to contradictions and hence is not compatible with any type of stochastic motion.

1·2. *The Poisson–Markoff or emigration-immigration process.* For setting up this process, we start with the following assumptions:

(1) The particles are moving independently of each other.

(2) The probability that precisely one of the particles in the region R_r at time t moves into R_s in time dt is $N_{t,r}\lambda_{rs}dt+o(dt)$.

(3) The probability that precisely one of the particles in R_r at time t moves into R^* in time dt is $N_{t,r}\lambda_r^*dt+o(dt)$.

(4) The probability that precisely one of the particles in R^* at time t moves into R_r in time dt is $\mu_r dt+o(dt)$.

The process $\{\mathbf{N}_t\}$ is then Markovian (of the first order) and the distribution of $\mathbf{N}_t$ at a single instant is, as pointed out below, that of k independent Poisson variates. This Poisson–Markoff process is also called the emigration-immigration process (Bartlett, 1949). The parameters λ_{rs}, λ_r^* and μ_r will remain unspecified for the moment. The detailed behaviour of the Poisson–Markoff process is discussed below.

Let (i) the 'stochastic interaction-rate' matrix (cf. Ruben & Rothschild, 1955)

$$\begin{bmatrix} \lambda_1^* + \sum_{r \neq 1} \lambda_{1r} & -\lambda_{12} & -\lambda_{13} & \dots & -\lambda_{1k} \\ -\lambda_{21} & \lambda_2^* + \sum_{r \neq 2} \lambda_{2r} & -\lambda_{23} & \dots & -\lambda_{2k} \\ \dots & \dots & \dots & \dots & \dots \\ -\lambda_{k1} & -\lambda_{k2} & -\lambda_{k3} & \dots & \lambda_k^* + \sum_{r \neq k} \lambda_{kr} \end{bmatrix} \tag{1}$$

be denoted by $\mathbf{\Lambda}$;

(ii) $\kappa_1, \dots, \kappa_k$ be the latent roots† of the matrix $\mathbf{\Lambda}$ and $\mathbf{K}$ the diagonal matrix with $\kappa_1, \dots, \kappa_k$ as elements; and

(iii) the latent row vector matrix of $\mathbf{\Lambda}$ be denoted by $\mathbf{\Theta}$.

Then in statistical equilibrium the probability generating function of $\mathbf{N}_{t+\tau}$ and $\mathbf{N}_t$ is given by

$$\Pi_{t,t+\tau}(\mathbf{z}_t, \mathbf{z}_{t+\tau}) \equiv \exp\{\mathbf{m}'(\mathbf{z}_t - \mathbf{1}) + \mathbf{m}'(\mathbf{z}_{t+\tau} - \mathbf{1}) + \mathbf{m}'(\mathbf{Z}_t - \mathbf{I})\,\mathbf{P}(\tau)\,(\mathbf{z}_{t+\tau} - \mathbf{1})\}, \tag{2}$$

where the meaning of the symbols in (2) is as follows. The column vector $\mathbf{m}$ is defined by $\boldsymbol{\mu}' = \mathbf{m}'\mathbf{\Lambda}$, $\boldsymbol{\mu}$ being the column vector with $\mu_1, \dots, \mu_k$ as elements; $(\mathbf{z}_T - \mathbf{1})$ $(T = t, t+\tau)$ is the vector with $z_{T,r} - 1$ $(r = 1, 2, \dots, k)$ as elements, $z_{T,r}$ being the auxiliary variable corresponding to $N_{T,r}$; $\mathbf{Z}_t - \mathbf{I}$ is the diagonal matrix with the elements $z_{t,r} - 1$ $(r = 1, \dots, k)$; and $\mathbf{P}(\tau) \equiv \mathbf{\Theta}^{-1} e^{-\mathbf{K}\tau} \mathbf{\Theta}$.

The proof of result (2), on the same lines as in Bartlett (1949), is given in Appendix 1. The marginal distribution of $\mathbf{N}_t$ is, from (2),

$$\Pi_t(\mathbf{z}_t) \equiv \exp\{\mathbf{m}'(\mathbf{z}_t - \mathbf{1})\}, \tag{3}$$

that is, that of k independent Poisson variates. As the process $\{\mathbf{N}_t\}$ is Markovian, the two-point probability distribution (2) of $\mathbf{N}_t$ is sufficient to specify the process fully.

1·3. *The kinematic hypothesis.* Nothing has been assumed about the form of the λ's and the μ's as yet. The following assumptions about them are now made, the assumptions being in agreement with the collision frequency considerations in the kinetic theory of gases. The proof of their agreement (cf. Rothschild, 1953*a*) is given in Appendix 2.

Let A_r be the area of the region R_r, B_{rs} be the length of the part of its boundary which it shares with R_s (where $r \neq s$) and L_r the 'free' boundary of R_r (which it shares with R^*). For some values of r, L_r may be zero; $\sum_{s \neq r} B_{rs} + L_r$ will always be the perimeter of R_r.

The kinematic hypothesis then asserts that

$$\text{(i) } \lambda_{rs} = \bar{c}\frac{B_{rs}}{\pi A_r}, \quad \text{(ii) } \lambda_r^* = \bar{c}\frac{L_r}{\pi A_r} \quad \text{and} \quad \text{(iii) } \mu_r = m_0 \rho_r L_r, \tag{4}$$

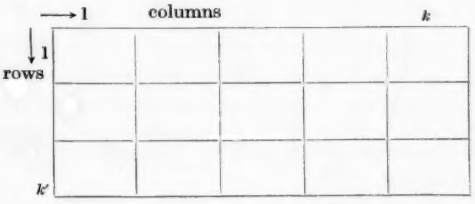
where the ρ_r are to be so chosen that the conditions $m_0 \mathbf{A}'\mathbf{\Lambda} = \boldsymbol{\mu}'$ is satisfied, $\bar{c}$ being the mean speed of the particles, m_0 the mean number of particles per unit area and $\mathbf{A}'$ the vector $(A_1, \dots, A_k)$. It is easy to verify that the ρ_r's thus chosen are all equal to $\bar{c}/\pi$.

† It is assumed that corresponding to a $r$ple latent root, there exist r linearly independent latent vectors, whatever matrix $\mathbf{\Lambda}$ is. This is true in the particular case considered in §1·5, as there $\mathbf{\Lambda}$ is a symmetric matrix.

The number of parameters are now reduced to two, namely, $\bar{c}$ and m_0. Thus the combination of the Poisson–Markoff process with the kinematic hypothesis leads to a two-parameter stochastic process, which has in fact been used in the study of the density fluctuations of spermatozoa and which we wish to examine here. ($\bar{c}$ and m_0 may be any two parameters and not necessarily the mean speed and the mean density of the particles, **A** being replaced by some other relevant vector and the discussion below still holds.)

1·4. There are two aspects of this Poisson–Markoff process to be investigated. First, is the Poisson–Markoff process when supposed to be valid for all modes of division of the space, consistent in the sense of being free from contradictions? If it is not, then evidently no type of stochastic motion can lead to it. Secondly, does the Poisson–Markoff process, irrespective of its theoretical appropriateness, give a 'good fit' to the data when the method of dividing up the space has been agreed upon beforehand and fixed? The second aspect is dealt with in relation to some data of spermatozoa counts in Part II of this paper. Regarding the first part, it is shown below that the postulation of the Poisson–Markoff process valid for all modes of division of the space is not consistent. As a simple case, consider the Poisson–Markoff model set up for $\mathbf{N}_t$ in three contiguous regions R_1, R_2 and R_3. By averaging over the region R_3, the 'marginal' process concerning R_1 and R_2 alone can be derived. Also just these two regions could have been considered at first and the Poisson–Markoff process for $\mathbf{N}_t$ in these two regions could have been set up. Then the 'consistency' principle would require that the two ways of finding the joint distribution of $N_{t,1}$ and $N_{t,2}$ should lead to the same answer. The consistency principle could be generalized to all the possible arrays of regions and selections from them. Here we consider the Poisson–Markoff process for $\mathbf{N}_t$ in a two-dimensional network of congruent rectangles, to be commonly met with in practice, and illustrate how the consistency principle breaks down.

1·5. We set up the Poisson–Markoff process with the kinematic hypothesis for $\mathbf{N}_t$ in this two-dimensional network of k' rows and k columns as shown in the accompanying figure and then consider the marginal process $\{N_t\}$ in one of the $k \times k'$ regions of the network:



We first specify the stochastic interaction-rate matrix $\mathbf{\Lambda}$ and the correlation matrix $\mathbf{P}(\tau)$ for this network of $k \times k'$ regions. For this case,

(1) $A_r = A$, say, for all r.

(2) $B_{rs} = \begin{cases} B, \text{ when } R_r \text{ and } R_s \text{ are contiguous regions in the same row,} \\ Bp', \text{ when } R_r \text{ and } R_s \text{ are in the same column,} \\ 0, \text{ otherwise.} \end{cases}$

(3) $\frac{1}{B}(L_r + \sum_{r \neq s} B_{rs}) = p$, for every region.

Obviously $p = 2(1+p')$.

Let $B\bar{c}/A\pi$ be denoted by λ. Then

$$\mathbf{\Lambda} \equiv \mathbf{\Lambda}_{kk'} = \begin{bmatrix} \mathbf{\Lambda}_k & -p'\lambda\mathbf{I}_k & (\mathbf{0}) & (\mathbf{0}) & \dots & \dots & (\mathbf{0}) \\ -p'\lambda\mathbf{I}_k & \mathbf{\Lambda}_k & -p'\lambda\mathbf{I}_k & (\mathbf{0}) & \dots & \dots & (\mathbf{0}) \\ 0 & -p'\lambda\mathbf{I}_k & \mathbf{\Lambda}_k & -p'\lambda\mathbf{I}_k & \dots & \dots & (\mathbf{0}) \\ 0 & 0 & -p'\lambda\mathbf{I}_k & \mathbf{\Lambda}_k & \dots & \dots & (\mathbf{0}) \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & -p'\lambda\mathbf{I}_k \\ 0 & 0 & 0 & 0 & \dots & -p'\lambda\mathbf{I}_k & \mathbf{\Lambda}_k \end{bmatrix}, \qquad (5)$$

where $\mathbf{\Lambda}_k$ is the $k \times k$ matrix

$$\lambda \begin{bmatrix} p & -1 & 0 & 0 & \dots & \dots & 0 \\ -1 & p & -1 & 0 & \dots & \dots & 0 \\ 0 & -1 & p & -1 & \dots & \dots & 0 \\ 0 & 0 & -1 & p & \dots & \dots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & -1 \\ 0 & 0 & 0 & 0 & 0 & -1 & p \end{bmatrix}$$

and $\mathbf{I}_k$ is the unit $k \times k$ matrix.

The matrix $\mathbf{\Lambda}_{kk'}$ may also be expressed as

$$\mathbf{\Lambda}_{kk'} = \mathbf{\Lambda}_k(\times)\,\mathbf{I}_{k'} + \mathbf{I}_k(\times)\,\mathbf{J}_{k'},$$

where the symbol $(\times)$ denotes the outer product, $\mathbf{I}_{k'}$ the unit $k' \times k'$ matrix and $\mathbf{J}_{k'}$ is the $k' \times k'$ matrix

$$\mathbf{J}_{k'} \equiv p'\lambda \begin{pmatrix} 0 & -1 & 0 & 0 & \dots & \dots & 0 \\ -1 & 0 & -1 & 0 & \dots & \dots & 0 \\ 0 & -1 & 0 & -1 & \dots & \dots & 0 \\ 0 & 0 & -1 & 0 & \dots & \dots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & -1 \\ 0 & 0 & 0 & 0 & 0 & -1 & 0 \end{pmatrix}.$$

For obtaining the latent roots and latent vectors of $\mathbf{\Lambda}_{kk'}$, we observe that

(1) the latent roots of $\mathbf{\Lambda}_k$ are

$$\kappa_r = p\lambda - 2\cos\omega_r \quad (r = 1, 2, \dots, k),$$

where $\omega_r = r\pi/(k+1)$;

(2) its latent row vector $\boldsymbol{\eta}'_r$ corresponding to κ_r is

$$\boldsymbol{\eta}'_r = \sqrt{\left(\frac{2}{k+1}\right)}(\sin k\omega_r, \sin(k-1)\omega_r, \dots, \sin\omega_r) \quad \text{(Montroll, 1947)};$$

(3) the latent roots of $\mathbf{J}_{k'}$ are

$$l_{r'} = -2\lambda p' \cos\omega'_{r'} \quad (r' = 1, 2, \dots, k),$$

where $\omega'_{r'} = \dfrac{r'\pi}{k'+1}$;

(4) the latent vectors of $\mathbf{J}_{k'}$ are the same as those of $\mathbf{\Lambda}_{k'}$. We denote the latent root matrices of $\mathbf{\Lambda}_k$ and $\mathbf{J}_{k'}$ by $\mathbf{K}_k$ and $\mathbf{L}_{k'}$ respectively and their latent vector matrices by $\mathbf{\Theta}_k$ and $\mathbf{\Theta}_{k'}$ respectively. Then the latent root matrix and the latent vector matrix of $\mathbf{\Lambda}_{kk'}$ are respectively given by

$$\left.\begin{aligned}\mathbf{K}_{kk'} &= \mathbf{K}_k(\times)\mathbf{I}_{k'} + \mathbf{I}_k(\times)\mathbf{L}_{k'},\\ \mathbf{\Theta}_{kk'} &= \mathbf{\Theta}_k(\times)\mathbf{\Theta}_{k'}.\end{aligned}\right\} \tag{6}$$

For
$$\begin{aligned}\mathbf{\Theta}_{kk'}\mathbf{\Lambda}_{kk'}\mathbf{\Theta}'_{kk'} &= \mathbf{\Theta}_k(\times)\mathbf{\Theta}_{k'}[\mathbf{\Lambda}_k(\times)\mathbf{I}_{k'} + \mathbf{I}_k(\times)\mathbf{J}_{k'}]\mathbf{\Theta}'_k(\times)\mathbf{\Theta}'_{k'}\\ &= [\mathbf{\Theta}_k\mathbf{\Lambda}_k\mathbf{\Theta}'_k](\times)[\mathbf{\Theta}_{k'}\mathbf{I}_{k'}\mathbf{\Theta}'_{k'}] + [\mathbf{\Theta}_k\mathbf{I}_k\mathbf{\Theta}'_k](\times)[\mathbf{\Theta}_{k'}\mathbf{J}_{k'}\mathbf{\Theta}'_{k'}]\\ &= \mathbf{K}_k(\times)\mathbf{I}_{k'} + \mathbf{I}_k(\times)\mathbf{L}_{k'},\end{aligned}$$

on our observing that $\mathbf{\Theta}'_{kk'} = \mathbf{\Theta}'_{k'}(\times)\mathbf{\Theta}'_{k'}$ and that as $\mathbf{\Lambda}_{k'}$, $\mathbf{J}_{k'}$ are symmetric matrices $\mathbf{\Theta}_k$ and $\mathbf{\Theta}_{k'}$ may be taken to be orthogonal matrices.

As the regions are equal in area, the m's are equal and the matrix

$$\mathbf{P}(\tau) \equiv \mathbf{\Theta}^{-1} e^{-\mathbf{K}\tau}\mathbf{\Theta} \equiv \mathbf{\Theta}' e^{-\mathbf{K}\tau}\mathbf{\Theta},$$

$\mathbf{\Theta}$ being an orthogonal matrix, reduces to the correlation matrix between $\mathbf{N}_t$ and $\mathbf{N}_{t+\tau}$. Making use of (6) it can be easily shown that (i) the auto-correlation coefficient for the region in the lth column and l'th row is

$$\begin{aligned}P_{ll',ll'}(\tau) = e^{-p\lambda\tau} &\times \frac{2}{k+1}\left\{\sum_{r=1}^{k}[\sin^2 l\omega_r]\, e^{2\lambda\tau\cos\omega_r}\right\}\\ &\times \frac{2}{k'+1}\left\{\sum_{r'=1}^{k'}[\sin^2 l'\omega'_{r'}]\, e^{2p'\lambda\tau\cos\omega'_{r'}}\right\},\end{aligned} \tag{7}$$

and (ii) the cross-correlation coefficient pertaining to the region in the l_1th column and the l'_1th row and the region in the l_2th column and the l'_2th row is

$$\begin{aligned}P_{l_1l'_1,l_2l'_2}(\tau) = e^{-p\lambda\tau} &\times \frac{2}{k+1}\left\{\sum_{r=1}^{k}[\sin l_1\omega_r][\sin l_2\omega_r]\, e^{2\lambda\tau\cos\omega_r}\right\}\\ &\times \frac{2}{k'+1}\left\{\sum_{r'=1}^{k'}[\sin l'_1\omega'_{r'}][\sin l'_2\omega'_{r'}]\, e^{2p'\lambda\tau\cos\omega'_{r'}}\right\}.\end{aligned} \tag{8}$$

The two-point probability distribution (2) for $\mathbf{N}_t$ in the network of $k \times k'$ regions is thus completely specified, and thus the process $\{\mathbf{N}_t\}$, being Markovian, may be fully specified. From this multivariate process, the process $\{N_t\}$ pertaining to the region in the lth column and l'th row may be obtained as a marginal process. In particular, the two-point probability distribution corresponding to this process is

$$\Pi_{l,l+\tau}(z_l, z_{l+\tau}) \equiv \exp\{m(z_l - 1) + m(z_{l+\tau} - 1) + mP_{ll',ll'}(\tau)(z_l - 1)(z_{l+\tau} - 1)\}. \tag{9}$$

The consistency principle requires that the distribution (9) should be invariant to changes in l, l', k and k'. An examination of the expression (7) for $P_{ll',ll'}(\tau)$ shows that this is not the case. Similarly, the consideration of other more complicated situations would be expected to lead to the same conclusion.

Formula (7) and hence the result (9) can easily be verified for simpler cases, say, for

(i) $k = k' = 1$, $l = l' = 1$ (i.e. the Poisson–Markoff model for region R_1 alone is considered),

and (ii) $k = 2$ and $k' = 1$, $l = l' = 1$ (i.e. the Poisson model for region R_1 is obtained as a marginal model from the Poisson–Markoff model for regions R_1 and R_2).

Then $$P(\tau) = e^{-p\lambda\tau}, \quad \text{in case (i)}, \tag{7a}$$

$$= e^{-p\lambda\tau}\cosh\lambda\tau, \quad \text{in case (ii)}. \tag{7b}$$

Since $\lambda \neq 0$, $e^{-p\lambda\tau} \neq e^{-p\lambda\tau}\cosh\lambda\tau$.

Hence the inconsistency of the Poisson–Markoff model is evident.

It follows, therefore, that there does not exist any stochastic motion that leads to the Poisson–Markoff process for $\mathbf{N}_t$, valid for all modes of division of the space. However, the kinematic hypothesis ensures, that for small τ, the Poisson–Markoff process is consistent to $O(\tau)$. Also in certain situations at least, the Poisson–Markoff process might be an adequate approximation of a consistent process for $\mathbf{N}_t$ governing any particular data of density fluctuations. This contention involves the question of goodness of fit and is examined in Part II of this paper, in relation to some actual experimental data.

Part II

2·1. *Goodness of fit tests.* The multivariate Poisson–Markoff model for $\mathbf{N}_t$, for large mean number density, degenerates into a normal Markoff model. Hence it can be represented asymptotically as a multivariate autoregressive model of the first order and the asymptotic χ^2-goodness of fit test based on serial correlations (the G-test or the H-test) (Bartlett & Rajalakshman, 1953; Quenouille, 1947) can then be applied.

Thus denoting the conditional expectation of $\mathbf{N}_t$ for given $\mathbf{N}_0$ at time t_0, $t \geqslant t_0$ by $E(\mathbf{N}_t \mid \mathbf{N}_0)$, we have from (2), § (1·2),

$$E(\mathbf{N}_t - \mathbf{m} \mid \mathbf{N}_0) = \mathbf{P}(t-t_0)'\,(\mathbf{N}_0 - \mathbf{m}), \tag{1}$$

a linear regression formula. Also

$$E[(\mathbf{N}_t - \mathbf{m})\,(\mathbf{N}_{t+\tau} - \mathbf{m})'] = \mathbf{M}\mathbf{P}(\tau), \tag{2}$$

$\mathbf{M}$ being the diagonal matrix $(m_1, m_2, \ldots, m_k)$. Therefore, the model may be specified by the stochastic equations

$$(\mathbf{N}_t - \mathbf{m}) - \mathbf{P}(t-t_0)'\,(\mathbf{N}_0 - \mathbf{m}) = \boldsymbol{\epsilon}_{t_0,t}, \tag{3}$$

where the probability distribution of $\boldsymbol{\epsilon}_{t_0,t}$ may be written down from that of the process $\{\mathbf{N}_t\}$. In particular, when $t_0 = t-\tau$, τ being fixed and t takes the values $a + r\tau$ ($r = 0, \pm, \pm 2, \ldots$)

$$E(\boldsymbol{\epsilon}_t) = (\mathbf{0}) \quad \text{and} \quad E(\boldsymbol{\epsilon}_t \boldsymbol{\epsilon}_{t'}') = (\mathbf{0}) \quad (t \neq t').$$

As $m_0 \to \infty$, the Poisson–Markoff model tends to a normal Markoff model and the variates $\boldsymbol{\epsilon}_t$ tend to be normal and to be independently distributed of $\boldsymbol{\epsilon}_{t'}$ ($t \neq t'$). In this case, the asymptotic H-test or the G-test can be used to assess the goodness of fit of the model (3) to any data consisting of a large number of observations. For finite m_0, $\boldsymbol{\epsilon}_t$ and $\boldsymbol{\epsilon}_{t'}$ ($t \neq t'$) are not distributed independently, as an examination of the joint distribution of $\boldsymbol{\epsilon}_t$ and $\boldsymbol{\epsilon}_{t'}$ would show. The effect of this dependence on the G-test or the H-test can, however, be assessed. This is done in § 2·2. In § 2·3 is given the summary of the asymptotic χ^2-goodness of fit test applied to bull spermatozoa data.

2·2. *The effect of the dependence of the residuals $\boldsymbol{\epsilon}_t$ for finite m_0 on the G-test.* We consider the one-variate case first. From equation (3) we have

$$(N_t - m) - \rho(N_{t-1} - m) = \epsilon_t,$$

the interval of time between consecutive observations being taken as the unit time and $\rho \equiv P(1) = e^{-\lambda}$, where the probability of one particle leaving R in time dt is $\lambda\,dt + o(dt)$.

This can be written as

$$H_t X_t = \epsilon_t, \tag{4}$$

where $H_t = 1 - \rho E_t^{-1}$, E_t^{-1} being the displacement operator changing X_t to X_{t-1}, and $X_t \equiv N_t - m$. Then since $E(X_{t+u}X_t) = \rho^{|u|}m$, it follows that

$$E(\epsilon_{t+u}X_t) = 0 \quad (u>0) \qquad \text{and} \qquad E(\epsilon_{t+u}\epsilon_{t+u'}) = m(1-\rho^2)\,\delta(u-u'), \tag{5}$$

where $$\delta(u-u') = 0, \quad (u \neq u'); \quad = 1, \quad (u = u').$$

Let $D_t \equiv H_t C_t$, where $C_t = \dfrac{1}{n-t}\sum_{u=1}^{n-t} X_u X_{t+u}$. Then

$$D_t \sim \frac{1}{n-t}\sum_{u=1}^{n-t} X_u \epsilon_{t+u},$$

$$\left.\begin{aligned} E(D_t) &\sim \frac{1}{n-t}E\left\{\sum_{u=1}^{n-t} X_u \epsilon_{t+u}\right\} = 0 \quad (t>0)\\ \text{and} \qquad E(H_t D_t) &= E(D_t - \rho D_{t-1}) \sim 0 \quad (t>1). \end{aligned}\right\} \tag{6}$$

Further it can be shown that

$$E(D_t D_\tau) \sim \frac{m^2\rho^{t-\tau}\,m^2(1-\rho^2)}{n-\tau} + \frac{m(1-\rho)\rho^t}{n-\tau} \quad (t \geqslant \tau). \tag{7}$$

Hence from the result (7)

$$\begin{aligned} E(H_t D_t H_\tau D_\tau) &= E[(D_t - \rho D_{t-1})(D_\tau - \rho D_{\tau-1})]\\ &= E(D_t D_\tau) - \rho E(D_t D_{\tau-1}) - \rho E(D_{t-1} D_\tau) + \rho^2 E(D_{t-1} D_{\tau-1})\\ &\sim \begin{cases} 0 & (t \neq \tau)\\ \dfrac{m(1-\rho)^2\rho^t}{n-t} + m^2\dfrac{(1-\rho^2)^2}{n-t} & (t=\tau). \end{cases} \end{aligned} \tag{8}$$

If $r_t \equiv C_t/C_0$, then

$$E(H_t^2 r_t H_\tau^2 r_\tau) \sim \left\{\begin{aligned} &0 && (t \neq \tau),\\ &\frac{(1-\rho)^2\rho^t}{m(n-t)} + \frac{(1-\rho^2)^2}{n-t} && (t=\tau). \end{aligned}\right\} \tag{9}$$

The effect of the residuals ϵ_t (cf. equation (14)) is, therefore, only to raise the variance of $H_t^2 r_t$ since $\rho > 0$ here, while $H_t^2 r_t$ and $H_\tau^2 r_\tau$ are still asymptotically uncorrelated when $t \neq \tau$. The proportionate increase in the variance of $H_t^2 r_t$ is less than $1/m$ and hence tends to zero, as it must do, when $m \to \infty$.

Working on the same lines for the Poisson–Markoff model for $\mathbf{N}_t$ in two contiguous equal regions, we get results similar to those in the one-variate case, namely, only the variances of the Q's (the linear combinations of the serial correlation coefficients for the G-test) are increased and the simultaneous covariances of the Q's are altered, both the proportionate increments being less than $1/m_0$. This indicates that similar results may be expected for the k-variate Poisson–Markoff model for $\mathbf{N}_t$ in general.

Similarly, the effect of the dependence of the residuals on the H-test can be determined and shown to be negligible for large m_0.

2·3. *Results of the G-test applied to spermatozoa data.* The data consist of the series of observed number of spermatozoa in 20 equal squares† forming a $4h \times 5h$ rectangle, observed

† Actually the twenty subregions are rectangles $63{\cdot}4 \times 67{\cdot}8\mu$.

at regular intervals of time, in film 186 of Lord Rothschild's data of bull spermatozoa. (For detailed description of the data see Rothschild (1953*b*).) For the purpose of the test, the twenty squares are divided into four groups of five squares each forming four rectangles in a line and the observations in the five squares of each of the groups are pooled. The interval of time between consecutive observations is taken as the unit of time.

For this case,

$$\mathbf{\Lambda}_4 = \lambda \begin{pmatrix} \frac{12}{5} & -1 & 0 & 0 \\ -1 & \frac{12}{5} & -1 & 0 \\ 0 & -1 & \frac{12}{5} & -1 \\ 0 & 0 & -1 & \frac{12}{5} \end{pmatrix}.$$

Equation (3) of § 2·1, viz.

$$(\mathbf{N}_t - \mathbf{m}) - \mathbf{P}(1)' (\mathbf{N}_{t-1} - \mathbf{m}) = \boldsymbol{\epsilon}_t, \tag{10}$$

where $\boldsymbol{\epsilon}_t$ is for large $\mathbf{m}$, a normal residual vector distributed independently of $\boldsymbol{\epsilon}_{t'}$, and $\mathbf{P}(1) = \mathbf{\Theta}_4' e^{-\mathbf{K}_4} \mathbf{\Theta}_4$, where $\mathbf{\Theta}_4$ and $\mathbf{K}_4$ are as defined in § 1·2, is transformed to

$$\mathbf{A}_t - e^{-\mathbf{K}_4} \mathbf{A}_{t-1} = \boldsymbol{\eta}_t, \tag{11}$$

where $\mathbf{\Theta}_4(\mathbf{N}_t - \mathbf{m}) = \mathbf{A}_t$ and $\boldsymbol{\eta}_t = \mathbf{\Theta}_4 \boldsymbol{\epsilon}_t$. $\boldsymbol{\eta}_t$ and $\boldsymbol{\eta}_{t'}$ $(t \neq t')$ are now asymptotically independently distributed. Let the vector $\mathbf{A}_t$ have A_t, B_t, C_t and D_t as its elements. The correlation coefficients between A_t, $A_{t+\tau}$ and A_t, $B_{t+\tau}$ are respectively denoted by $\rho_{AA}(\tau)$, $\rho_{AB}(\tau)$ and similar notations are used for the other auto- and cross-correlations. It can easily be seen that the cross-correlations are zero and $\rho_{AA}(\tau) = e^{-\kappa_1 \tau}$. The corresponding serial auto- and cross-correlations are denoted by $r_{AA}(\tau)$, $r_{AB}(\tau)$, etc., and the expressions used for them are, e.g.

$$r_{AB}(\tau) \equiv \frac{\sum_{t=1}^{n-\tau} (a_t - \bar{a})(b_{t+\tau} - \bar{b})/(n-\tau)}{\sqrt{\left[\sum_{t=1}^{n} (a_t - \bar{a})^2/n \sum_{t=1}^{n} (b_t - \bar{b})^2/n\right]}},$$

where a_t and b_t are the observed values corresponding to A_t and B_t respectively,

$$\bar{a} = \sum_{t=1}^{n} a_t/n \quad \text{and} \quad \bar{b} = \sum_{t=1}^{n} b_t/n.$$

For the data, $n = 100$. The χ^2's for the G-test are

$$\chi_1^2 = \sum_{l \text{ and } m} \left[\frac{(r_{lm}(1) - r_{lm}(0)\rho_{mm}(1))^2}{(1-\rho_{mm}^2(1))/n}\right]$$

and

$$\chi_t^2 = \sum_{l \text{ and } m} \left[\frac{(r_{lm}(t) - r_{lm}(t-1)(\rho_{ll}(1) + \rho_{mm}(1)) + r_{lm}(t-2)\rho_{ll}(1)\rho_{mm}(1)^2}{(1-\rho_{ll}^2(1))(1-\rho_{mm}^2(1))/n}\right] \quad (t > 1),$$

where the summations are over l and m, l and m becoming A, B, C and D.

The observed mean densities for the four rectangles are 32·17, 28·94, 31·35, 36·37, the interval of time between consecutive counts is 0·4 sec. and $h \sim 65{\cdot}5\mu$.

The values of χ_1^2 and χ_2^2 are given in the first row of Table 1, the estimate used for $\rho \equiv e^{-\lambda}$ being

$$r = [\Sigma\{\lambda_A/\text{var}(\lambda_A)\}/\Sigma\{1/\text{var}(\lambda_A)\}]_{\rho=r}, \tag{12}$$

where the summation is over A, B, C and D,

$$\lambda_A = (r_{AA}(1))^{s_A}$$

and $s_A^{-1} = \frac{12}{5} - 2\cos\omega_1$, $s_B^{-1} = \frac{12}{5} - 2\cos\omega_2$, $s_C^{-1} = \frac{12}{5} - 2\cos\omega_3$, $s_D^{-1} = \frac{12}{5} - 2\cos\omega_4$, ω_r being equal to $r\pi/5$. It may be noted that $s_A = \lambda/\kappa_1$, $s_B = \lambda/\kappa_2$, etc.

Some empirical values of ρ were also used to calculate χ_1^2 and χ_2^2 and in rows 2 and 3 in Table 1 are given the values of χ_1^2 and χ_2^2 for those values of ρ for which $\chi_1^2+\chi_2^2$ appeared to be minimum. (For individual correlations see Table 4 given in Appendix 3.) The χ^2's are highly significant, indicating a departure of the data from the four-variate Poisson–Markoff model.

Table 1

Film 186	χ_1^2	χ_2^2	$\chi_1^2+\chi_2^2$
$\hat{\rho}=r=0{\cdot}9208$ as obtained from equation (12)	76	310	386
D.F.	15	16	31
$\rho=0{\cdot}82$ for which entry in col. 3 appears	105	71	176
$=0{\cdot}84$ to be minimum	94	81	175
D.F.	15	16	31

The G-test in the one-variate case. The data used for the one-variate case are the five independent series of observations of spermatozoa counts in squares (film 191, shot 1, shot 3, shot 4, shot 5, shot 6 of Lord Rothschild's data of bull spermatozoa). The G-test is applied to each of the series separately. m is estimated by the sample mean and ρ, either by (i) r, the serial correlation coefficient, or by (ii)

$$r' \equiv 1 - \frac{\sum_{t=2}^{n}(n_t - n_{t-1})^2/(n-1)}{2\sum_{t=1}^{n} n_t/n}.$$

It should be noticed that the variances of r and r' are

$$\text{var}(r) \sim \frac{1}{n}(1-\rho^2) + \rho\left(\frac{1-\rho}{mn}\right)$$

and

$$\text{var}(r') \sim \frac{1}{n}\frac{(3+\rho)(1-\rho)^2}{1+\rho} + \frac{\rho(1-\rho)}{mn}$$

(Lindley, 1954; Ruben & Rothschild, 1955). Hence when $\rho > \sqrt{2}-1$, $\text{var}(r') < \text{var}(r)$ and therefore r' is a more efficient estimate of ρ than r.

The values of

$$\chi^2 \equiv \sum_{t=2}^{6}\chi^2(t) \equiv \sum_{t=2}^{6}(H_t^2 r_t)^2 \Big/ \frac{(1-\rho^2)^2}{n}\dagger$$

are given in Table 2 for the two cases (1) and (2). When r is used as the estimate of ρ, the χ^2-test does not indicate a significant departure of the data from the Poisson–Markoff model, whereas it does so when r' is used as the estimate of ρ. This should not be the case if the Poisson–Markoff hypothesis is true, since the errors of both the estimates r and r' are $O(n^{-\frac{1}{2}})$. On the other hand, if the Poisson–Markoff hypothesis is not valid, whether the discrepancy is brought out or not may depend upon the estimate used. To examine this question we may test whether the sample variance/mean ratio, $(C_0/\bar{n})$, which is asymptotically

† In both the uniregional and the multiregional cases the m's are fairly large and hence the effect of finite m on the tests has been neglected.

normally distributed, differs significantly from unity, the theoretical value assumed in the use of r'. On the hypothesis of the Poisson–Markoff model

$$\left.\begin{aligned} E\left(\frac{C_0}{\bar{n}}\right) &\sim 1-\frac{1}{n}\frac{1+\rho}{1-\rho} \\ \text{and} \qquad \text{var}\left(\frac{C_0}{\bar{n}}\right) &\sim \frac{2}{n}\left(\frac{1+\rho^2}{1-\rho^2}\right). \end{aligned}\right\} \tag{13}$$

Table 2

Film 191	Shot 1	Shot 3	Shot 4	Shot 5	Shot 6	Total
Estimate of $\rho = r$	0·5426	0·3755	0·5606	0·4038	0·4313	
χ^2	4·1	1·6	1·1	4·1	3·3	14·2
Estimate of $\rho = r'$	0·7644	0·6373	0·6596	0·6943	0·7714	
χ^2	18·7	6·4	1·9	16·2	39·1	82·3
D.F.	5	5	5	5	5	25
$m \equiv$ sample mean	24·9911	16·0446	15·8850	12·8889	20·7719	
$n \equiv$ no. of observations	112	112	113	117	114	
$\tau \equiv$ interval of time between consecutive observations in sec.	0·2279	0·2288	0·2250	0·2205	0·2248	

Table 3

Film 191	Shot 1	Shot 3	Shot 4	Shot 5	Shot 6	$\sqrt{5}\,\overline{C}$	t_C
C	−1·3	−1·7	−0·6	−1·8	−1·9	−3·3	−6·2

$\overline{C} \equiv$ A.M. of the C's for the five series. $t_C \equiv \sqrt{5}\,\overline{C}/\sqrt{[\Sigma(C-\overline{C})^2/4]}$.

This follows from

$$E(\bar{n}) = m, \quad \text{var}\,(\bar{n}) \sim \frac{m}{n}\left(\frac{1+\rho}{1-\rho}\right),$$

$$E(C_0) \sim m\left(1-\frac{1}{n}\frac{1+\rho}{1-\rho}\right), \quad \text{var}\,(C_0) \sim \frac{2m^2}{n}\frac{1+\rho^2}{1-\rho^2}+\frac{m}{n}\frac{1+\rho}{1-\rho},$$

$$\text{Cov}\,(C_0, \bar{n}) \sim \frac{m}{n}\frac{1+\rho}{1-\rho},$$

and using the method of statistical differentials to obtain $E(C_0/\bar{n})$ and $\text{var}\,(C_0/\bar{n})$. Table 3 gives the values of

$$C = \frac{C_0/\bar{n} - E(C_0/\bar{n})}{\sqrt{\text{var}\,(C_0/\bar{n})}} \sim \left(\frac{C_0}{\bar{n}} - 1 + \frac{1}{n}\frac{1+r'}{1-r'}\right) \Big/ \left(\frac{2}{n}\frac{1+r'^2}{1-r'^2}\right)^{\frac{1}{2}}$$

for the five series. The variate $\sqrt{5}\,\overline{C}$ is asymptotically normal with mean zero and unit variance, while t_C is a t-variate with 4 d.f. Both are here significantly large, bringing out

a significant deviation of the data from the one-variate Poisson–Markoff model with particular reference to the variance mean ratio.

The Poisson–Markoff model, therefore, does not appear to give a good fit either to the multiregional or to the uniregional bull spermatozoa data considered here. In the uniregional case the evidence for the departure of the data from the Poisson–Markoff model is not as striking as in the multiregional case, since the value of the pooled χ^2's for the five series is not significantly large when the serial correlation coefficient is used as an estimate of ρ.†

Since the Poisson–Markoff model is consistent only over small total intervals of time, it is desirable to specify the stochastic motion fully and then to base the stochastic models for $\mathbf{N}_t$ on it so that these models would be free from contradictions. The stochastic models for $\mathbf{N}_t$ are, in general, not Markovian (Bartlett, 1954). However, the closeness of approximation of the Poisson–Markoff model to such consistent models may then be studied, and the Poisson–Markoff model, which is comparatively simple, can be used as an approximate model in place of the original one, if the closeness of the approximation is found adequate for the particular situation under consideration. An investigation of some specific models for $\mathbf{N}_t$ based on fully specified stochastic motion has been carried out and it is hoped to publish an account of this work in due course.

I am extremely grateful to Prof. M. S. Bartlett for his valuable guidance. I am indebted to Lord Rothschild for making the spermatozoa data available to me; to Lord Rothschild and Dr H. Ruben and to Mr D. V. Lindley for allowing me to read the draft of their papers before their publication. I thank the referees for detailed suggestions regarding the first and second parts.

APPENDIX 1

Proof of formula (2), § 1·3

The infinitesimal probabilities of transition from one state to the others of the process $\{\mathbf{N}_t\}$ are expressed in tabular form below. Here we have written n_r for $N_{t,r}$.

Transition from state at time t to state at time $t+dt$	Probability of the transition
$n_r \to n_r+1$	$\mu_r dt + o(dt)$
$n_r \to n_r-1$	$n_r\lambda_r^* dt + o(dt)$
$\begin{cases} n_r \to n_r+1 \\ n_s \to n_s-1 \end{cases}$	$n_s\lambda_{sr} dt + o(dt)$
all other transitions from one state to a *different* state	$o(dt)$
Hence no change	$1-\sum_r \mu_r dt - \sum_r (\lambda_r^* + \sum_{s+r} \lambda_{rs})\, n_r dt - o(dt).$

Then the partial differential equation for $\Pi \equiv \Pi_t(z)$, the p.g.f. of $\mathbf{N}_t$, obtained as in Bartlett (1949) by taking into consideration all these possible transitions is

$$\frac{\partial \Pi}{\partial t} = \Pi \sum_r \mu_r(z_r - 1) + \sum_r \frac{\partial \Pi}{\partial z_r}\{\sum_{s+r} \lambda_{rs}(z_s - z_r) + \lambda_r^*(1-z_r)\}. \tag{1}$$

† The results for this particular test, which was made first, were quoted by Prof. Bartlett in his book *An Introduction to Stochastic Processes* (1955, p. 274); the other results had not then been obtained.

The auxiliary equations of this are

$$\frac{dt}{1} = -\frac{dz_r}{\sum_{s\neq r}\lambda_{rs}(z_s - z_r) + \lambda_r^*(1-z_r)} = \frac{d\Pi}{\Pi\sum_r \mu_r(z_r-1)}. \tag{2}$$

With the matrix $\mathbf{\Lambda}$ as given in § 1·2, let the vector $\mathbf{m}' \equiv (m_1, m_2, \ldots, m_k)$ be defined by

$$\boldsymbol{\mu}' = \mathbf{m}'\mathbf{\Lambda}. \tag{3}$$

(It should be noted that $|\mathbf{\Lambda}| \equiv \kappa_1\kappa_2\ldots\kappa_k \neq 0$; see line 4 from bottom of page.)

For any linearly transformed variate $\boldsymbol{\alpha}'(\mathbf{z}-\mathbf{1})$, equations (2) give

$$\frac{d\Pi}{\Pi\boldsymbol{\mu}'(\mathbf{z}-\mathbf{1})} = \frac{-d\boldsymbol{\alpha}'(\mathbf{z}-\mathbf{1})}{-\boldsymbol{\alpha}'\mathbf{\Lambda}(\mathbf{z}-\mathbf{1})} \tag{4}$$

and

$$\frac{dt}{1} = \frac{-d\boldsymbol{\alpha}'(\mathbf{z}-\mathbf{1})}{-\boldsymbol{\alpha}'\mathbf{\Lambda}(\mathbf{z}-\mathbf{1})}. \tag{5}$$

First take $\boldsymbol{\alpha}$ in (4) as identical with $\mathbf{m}$, and then, with the use of (3), we get

$$\frac{d\Pi}{\Pi} = \frac{d\mathbf{m}'(\mathbf{z}-\mathbf{1})}{1} \tag{6}$$

so that $\Pi\, e^{-\mathbf{m}'(\mathbf{z}-\mathbf{1})} = K_1$, where K_1 is any arbitrary constant. Next take $\boldsymbol{\alpha}'$ as identical with $\boldsymbol{\eta}_r'$, the latent row vector of $\mathbf{\Lambda}$ corresponding to the latent root κ_r, and equation (5) becomes

$$\frac{dt}{1} = \frac{-d\boldsymbol{\eta}_r'(\mathbf{z}-\mathbf{1})}{-\kappa_r\boldsymbol{\eta}_r'(\mathbf{z}-\mathbf{1})} \quad (r = 1, \ldots, k), \tag{7}$$

whence $\boldsymbol{\eta}_r'(\mathbf{z}-\mathbf{1})\, e^{-\kappa_r t} = \text{constant}$.

The general solution of equation (1) is therefore

$$\Pi_t(\mathbf{z})\, e^{-\mathbf{m}'(\mathbf{z}-\mathbf{1})} = \psi(\boldsymbol{\eta}_1'(\mathbf{z}-\mathbf{1})\, e^{-\kappa_1 t}, \quad \boldsymbol{\eta}_2'(\mathbf{z}-\mathbf{1})\, e^{-\kappa_2 t}, \quad \ldots, \quad \boldsymbol{\eta}_k'(\mathbf{z}-\mathbf{1})\, e^{-\kappa_k t}), \tag{8}$$

where ψ is an arbitrary function, the form of which may be determined from the initial conditions. If initially the numbers of particles in the regions $R_1, \ldots, R_k$ are $n_1, \ldots, n_k$ respectively, then taking the initial instant to correspond to $t = 0$,

$$\Pi_0(\mathbf{z}) \equiv \prod_{r=1}^{k} z_r^{n_r}.$$

Hence, putting $t = 0$ in (8),

$$\prod_{r=1}^{k} z_r^{n_r}\, e^{-\mathbf{m}'(\mathbf{z}-\mathbf{1})} = \psi(\boldsymbol{\eta}_1'(\mathbf{z}-\mathbf{1}), \ldots, \boldsymbol{\eta}_k'(\mathbf{z}-\mathbf{1})). \tag{9}$$

Now with $\mathbf{\Theta}$ defined as in § 1·2 let the vector $\mathbf{y}$ be defined by

$$\mathbf{\Theta}(\mathbf{z}-\mathbf{1}) = \mathbf{y} \quad \text{or equivalently} \quad \mathbf{z}-\mathbf{1} = \mathbf{\Theta}^{-1}\mathbf{y} \tag{10}$$

and let $\boldsymbol{\xi}_r'$, the row vector of $\mathbf{\Theta}^{-1}$ corresponding to $(z_r - 1)$. Then equation (9) can be written as

$$\psi(y_1, \ldots, y_k) = e^{-\sum_r (\boldsymbol{\xi}_r'\mathbf{y})m_r} \prod_{r=1}^{k} (1 + \boldsymbol{\xi}_r'\mathbf{y})^{n_r}. \tag{11}$$

With this expression for ψ, equation (8) becomes

$$\Pi_t(\mathbf{z}\,|\,\mathbf{n})\, e^{-\mathbf{m}'(\mathbf{z}-\mathbf{1})} = \exp\{-\sum_s\sum_r m_r[\xi_{rs}(\boldsymbol{\eta}_s'(\mathbf{z}-\mathbf{1})\, e^{-\kappa_s t})]\} \prod_{r=1}^{k} [1 + \sum_s \xi_{rs}(\boldsymbol{\eta}_s'(\mathbf{z}-\mathbf{1})\, e^{-\kappa_s t})]^{n_r}$$

$$= \exp\{-[\mathbf{m}'\mathbf{\Theta}^{-1}\, e^{-\mathbf{K}t}\mathbf{\Theta}(\mathbf{z}-\mathbf{1})]\} \prod_{r=1}^{k} [1 + \boldsymbol{\zeta}_r'(\mathbf{z}-\mathbf{1})]^{n_r}, \tag{12}$$

where $\boldsymbol{\zeta}_r'$ is the rth row vector of the matrix

$$\mathbf{P}(t) \equiv \mathbf{\Theta}^{-1}\, e^{-\mathbf{K}t}\mathbf{\Theta}.$$

It can be shown by expanding $\mathbf{\Lambda}$ in terms of λ_r^*'s in the diagonal elements and employing the method of mathematical induction that all the principal minors of $\mathbf{\Lambda}$ are positive-definite. Hence all the latent roots of $\mathbf{\Lambda}$ are positive (cf. Bartlett, 1949). Therefore when $t \to \infty$, then

$$e^{-\kappa_s t} \to 0, \quad \text{for all } s,$$

and (12) gives immediately

$$\Pi_\infty(\mathbf{z}\,|\,\mathbf{n}) = e^{\mathbf{m}'(\mathbf{z}-\mathbf{1})}. \tag{13}$$

This probability distribution is independent of the initial value of $\mathbf{N}_t$, i.e. value of $\mathbf{N}_t$ at $t = 0$. Hence for all t and in particular for $t = 0$ we take,

$$\Pi_t(\mathbf{z}) = e^{\mathbf{m}'(\mathbf{z}-\mathbf{1})}. \tag{14}$$

From formulae (12) and (14), expression (2) § 1·2 for the p.g.f. of $\mathbf{N}_t$ and $\mathbf{N}_{t+\tau}$ may be obtained.

APPENDIX 2

The proof that in two dimensions the collision frequency is $m_0\bar{c}/\pi$; also the proof of formulae (4) § 1·3

Let $\mathbf{c}$ be the vector $\begin{pmatrix} c_x \\ c_y \end{pmatrix}$ representing the velocity of a particle in two dimensions, and $f(\mathbf{c})$, the probability density for the velocity. Consider the number N of particles passing in time dt from one side of an infinitesimal element of length ds to the other. Then it is easily seen that

$$N = \int_{c_y>0} m_0 ds\, c_y dt f(\mathbf{c})\, d\mathbf{c}, \tag{1}$$

where the y-axis is taken along the normal to ds, the positive direction pointing towards that side of ds to which the particles move. On transforming to polar coordinates c and θ given by $c_x = c\cos\theta$, $c_y = c\sin\theta$, and integrating over θ, equation (1) becomes

$$N = \frac{m_0 ds dt}{\pi}\int_0^\infty c f(c)\, dc = \frac{m_0\bar{c}}{\pi}\, ds dt.$$

Hence the number of particles crossing a unit length in unit time is given by

$$\frac{m_0\bar{c}}{\pi}. \tag{2}$$

Therefore
$$m_r P_{rs}(dt) = \frac{m_0\bar{c}}{\pi} B_{rs} dt.$$

But with the assumption of a uniform mean density m_0,

$$m_r = m_0 A_r. \tag{3}$$

Hence
$$P_{rs}(dt) = \frac{\bar{c}B_{rs}}{\pi A_r} dt. \tag{4}$$

Similarly
$$P^*_r(dt) = \frac{\bar{c}L_r}{\pi A_r} dt. \tag{5}$$

Noting that in a state of macroscopic equilibrium the mean number of particles coming into a region R_r from R^* and the other regions in time dt balance the mean number of particles going out of R_r to R^* and the other regions in the same time, and using the equations (3), we obtain the relation

$$\boldsymbol{\mu}' = \mathbf{m}'\boldsymbol{\Lambda}. \tag{6}$$

APPENDIX 3

Tables of serial and cross-correlations

Table 4. *Values of* $r(\tau)$ for $\tau = 0, 1, 2$, *for the series* $\mathbf{A}_t$ (*defined in* § 2·3)

Film 186	$r_{(0)}$	$r_{(1)}$	$r_{(2)}$	Film 186	$r_{(0)}$	$r_{(1)}$	$r_{(2)}$
r_{AA}	1·0000	0·7207	0·6517	r_{CA}	−0·2688	−0·2506	−0·2551
r_{AB}	−0·5220	−0·5124	−0·4939	r_{CB}	0·4445	0·4381	0·4457
r_{AC}	−0·2688	−0·3259	−0·2836	r_{CC}	1·0000	0·8914	0·8461
r_{AD}	−0·4906	−0·4987	−0·4619	r_{CD}	0·2682	0·2617	0·2864
r_{BA}	−0·5220	−0·4846	−0·5877	r_{DA}	−0·4906	−0·5356	−0·5269
r_{BB}	1·0000	0·8265	0·7316	r_{DB}	0·5538	0·5508	0·5279
r_{BC}	0·4445	0·4486	0·4509	r_{DC}	0·2682	0·2389	0·2319
r_{BD}	0·5538	0·5801	0·5453	r_{DD}	1·0000	0·8725	0·8102

Table 5. *Values of r_τ, for $\tau = 1, 2, \ldots, 6$, for the five series of observations (Film 191, shot 1, shot 3, shot 4, shot 5, shot 6 of Rothschild's data)*

	r_1	r_2	r_3	r_4	r_5	r_6
Shot 1	0·5426	0·3109	0·1976	0·0588	0·0560	−0·0608
Shot 3	0·3755	0·1013	0·0784	0·1050	0·0258	0·0206
Shot 4	0·5606	0·3280	0·2026	0·1806	0·1235	0·1096
Shot 5	0·4038	0·2055	−0·0041	−0·1248	−0·1107	−0·0043
Shot 6	0·4313	0·1976	0·1617	−0·0033	−0·0520	0·0024

REFERENCES

Aitken, A. C. (1951). *Determinants and Matrices* (University Mathematical Texts). Edinburgh.

Bartlett, M. S. (1949). Some evolutionary stochastic processes. *J.R. Statist. Soc.* B, **11**, 211.

Bartlett, M. S. (1954). Processus stochastiques ponctuels. *Ann. Inst. Poincaré*, **14**, 35.

Bartlett, M. S. (1955). *An Introduction to Stochastic Processes.* Cambridge University Press.

Bartlett, M. S. & Rajalakshman, D. V. (1953). Goodness of fit tests for simultaneous autoregressive schemes. *J.R. Statist. Soc.* B, **15**, 107.

Chandrasekhar, S. (1943). Stochastic problems in Physics and Astronomy. *Rev. Mod. Phys.* **15**, 1.

Lindley, D. V. (1954). The estimation of the velocity distribution from counts. (Paper read at the Amsterdam Conference, 1954 and to be published in the Proceedings of the Conference.)

Montroll, E. W. (1947). On the theory of Markoff chains. *Ann. Math. Statist.* **18**, 18.

Quenouille, M. H. (1947). A large sample test for the goodness of fit of autoregressive schemes. *J.R. Statist. Soc.* **110**, 123.

Rothschild, Lord (1953*a*). A new method of measuring sperm speed. *Nature, Lond.*, **171**, 512.

Rothschild, Lord (1953*b*). A new method of measuring the activity of spermatozoa. *J. Exp. Biol.* **30**, 178.

Ruben, H. & Rothschild, Lord (1955). Estimation of mean speeds of organisms and particles by counting (to be published).

TESTING FOR SERIAL CORRELATION IN LEAST SQUARES REGRESSION

By E. J. HANNAN

Australian National University, Canberra, A.C.T.

1. Introduction

We shall be concerned with a regression of the form

$$\begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix} = \begin{bmatrix} x_{11} & x_{21} & \cdots & x_{k1} \\ x_{12} & x_{22} & \cdots & x_{k2} \\ \vdots & \vdots & \cdots & \vdots \\ x_{1n} & x_{2n} & \cdots & x_{kn} \end{bmatrix} \begin{pmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_k \end{pmatrix} + \begin{pmatrix} \epsilon_1 \\ \epsilon_2 \\ \vdots \\ \epsilon_n \end{pmatrix} \tag{1}$$

or

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon},$$

where the ϵ_t are generated by a stationary process, wholly independent of the vectors in $\mathbf{X}$, and have correlation matrix $\boldsymbol{\Gamma}$.

In a fundamental paper Anderson (1948) considered the case where the ϵ_t were generated by certain Gaussian processes giving a joint distribution near to that for a stationary simple Markoff process. For the cases he considered, when the column vectors of $\mathbf{X}$ are latent vectors of $\boldsymbol{\Gamma}$, a uniformly most powerful (one-sided) test of the hypothesis of serial independence of the ϵ_t may be obtained from the ratio of a quadratic form in the residuals to the sum of squares of the residuals. The matrix of the quadratic form in the numerator also has the vectors in $\mathbf{X}$ as latent vectors and is close to the matrix occurring in the definition of the serial correlation coefficient. His results suggest the use of some form of the serial correlation coefficient as a test statistic in the general case when the regressor vectors are no longer latent vectors of $\boldsymbol{\Gamma}$. Durbin & Watson (1950, 1951) use, for example, the statistic

$$d = \frac{\mathbf{z}'\mathbf{A}_d\mathbf{z}}{\mathbf{z}'\mathbf{z}},$$

where

$$\mathbf{z} = [\mathbf{I} - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}']\mathbf{y} = \mathbf{Q}\mathbf{y}$$

and

$$\mathbf{A}_d = \begin{bmatrix} 1 & -1 & & & & \\ -1 & 2 & -1 & & & \\ & -1 & 2 & \ddots & & \\ & & \ddots & \ddots & \ddots & \\ & & & \ddots & 2 & -1 \\ & & & & -1 & 1 \end{bmatrix}. \tag{2}$$

The elements of $\mathbf{A}_d$ outside the principal diagonal, and the diagonals immediately above and below it, are zeros.

Durbin & Watson reduced d to the canonical form

$$d = \frac{\sum_1^{n-k} \nu_j \zeta_j^2}{\sum_1^{n-k} \zeta^2},$$

where the ν_j are the latent roots of $\mathbf{QA}_d$ (other than k zeros) and the ζ_j are independent standard normal variates with zero mean when the ϵ_t are independent normal variates. They showed, moreover, that, ordering the ν_j with respect to size

$$\lambda_j \leqslant \nu_j \leqslant \lambda_{j+k-s},$$

where the λ_j are the latent roots of $\mathbf{A}_d$ (also ordered with respect to size) other than s which correspond to regressor vectors which are latent vectors of $\mathbf{A}_d$.

They then tabulated, for various n and k, the significance points of

$$d_l = \frac{\sum_1^{n-k} \lambda_j \zeta_j^2}{\sum_1^{n-k} \zeta_j^2} \quad \text{and} \quad d_u = \frac{\sum_1^{n-k} \lambda_{j+k-1} \zeta_j^2}{\sum_1^{n-k} \zeta_j^2}, \tag{3}$$

which correspond to the case where $s = 1$, and the only latent vector is that vector composed entirely of units, which corresponds to the constant term in the regression. These significance points then provide bounds to the significance point for d.

In general the ν_j depend upon the regressor vectors, the only simple case being, again, that where the regressors are latent vectors of $\mathbf{A}_d$. In cases where the bounds test is inconclusive they recommend the use of $\frac{1}{4}d$ as a Beta variate with the appropriate mean and variance.

An alternative test was suggested in Hannan (1955), but, while this test is asymptotically fully efficient, when n is small and k large relative to n end-effects reduce its power very greatly. In these circumstances the bounds test is also unsatisfactory however, since the bounds are very far apart. For example, with $n = 30$ and $k = 6$ the bounds are 1·07 and 1·83, although the range of variation of d is only from 0 to 4 (approximately).

As has already been seen, the latent vector case is of primary importance in connexion with the test for serial independence of the residuals. It is also of great importance in the estimation of $\boldsymbol{\beta}$, for when $\mathbf{X}$ is composed of latent vectors of $\boldsymbol{\Gamma}$ it is well known that the straightforward least squares (L.S.) estimates are, numerically, the same as the best linear unbiased (B.L.U.) estimates (obtained when $\boldsymbol{\Gamma}$ is known). Grenander (1954) and Grenander & Rosenblatt (1954) have investigated the conditions under which these two differing estimation procedures give, asymptotically, estimates having the same covariance matrix, no matter what the stationary process generating the ϵ_t may be. Their conditions, which will be given in more detail in the next section, are met by regressions on certain analytic functions (for equidistant values of the argument), including the orthogonal (Legendre) polynomials, the trigonometric functions and functions of the form $t^p \cos \theta t$, $t^p \sin \theta t$. The analogy with the latent vector case suggests that when Grenander & Rosenblatt's theorem applies the distribution theory of statistics such as d can be simplified and this is investigated in § 2.

Most attention will be paid to the orthogonal polynomials, since the regression on trigonometric functions has been dealt with exactly by Anderson & Anderson (1950) and other cases where Grenander & Rosenblatt's theorem holds seem less interesting. It should be mentioned that much of what is contained in § 3 (relating to the orthogonal polynomials) is already implicit in Durbin & Watson's work. (See, for example, Durbin & Watson (1951), pp. 171 and 172.)

In the final section of the paper the problem of testing for serial correlation in the residuals from a regression which includes trend terms as well as stochastic variables is considered.

2. Regressors whose spectral distribution function is a step function

Grenander (1954) and Grenander & Rosenblatt (1954) consider the case of regression on k vectors (of n elements each), $\boldsymbol{\phi}_1, \boldsymbol{\phi}_2, \ldots, \boldsymbol{\phi}_k$, generated by a process such that

$$\left.\begin{aligned} &\text{(i)} \quad \lim_{n\to\infty} \sum_{t=1}^{n} \phi_\mu^2(t) = \infty \quad (\mu = 1, \ldots, k), \\ &\text{(ii)} \quad \lim_{n\to\infty} \frac{\sum_{t=1}^{n+1} \phi_\mu^2(t)}{\sum_{t=1}^{n} \phi_\mu^2(t)} = 1 \quad (\mu = 1, \ldots, k), \\ &\text{(iii)} \quad \lim_{n\to\infty} \frac{\sum_{t=1}^{n} \phi_\mu(t+h)\,\phi_\nu(t)}{\left\{\sum_{t=1}^{n} \phi_\mu^2(t) \sum_{t=1}^{n} \phi_\nu^2(t)\right\}^{\frac{1}{2}}} = r_{\mu,\nu}(h), \text{ exists} \quad (\mu, \nu = 1, \ldots, k). \end{aligned}\right\} \tag{4}$$

The $\phi_\mu(t)$ are considered merely as known sequences of real numbers, no other restrictions on the process generating them being made.

Grenander & Rosenblatt then show that

$$\mathbf{R}_h = [r_{\mu,\nu}(h)] = \int_{-\pi}^{\pi} e^{ih\theta}\, d\mathbf{M}(\theta), \tag{5}$$

where $\mathbf{M}(\theta) = [m_{\mu,\nu}(\theta)]$ is a matrix of functions with $\mathbf{M}(\theta_2) - \mathbf{M}(\theta_1)$ (Hermitian) non-negative definite for $\theta_2 > \theta_1$. It is presumed that $\mathbf{R}_0 = \mathbf{M}(\pi) - \mathbf{M}(-\pi)$ is non-singular (to avoid a kind of asymptotic multicollinearity). $\mathbf{M}(\theta)$ is called the spectral distribution function of the $\boldsymbol{\phi}_\mu$. Since the $\phi_\mu(t)$ are real we shall have $d\mathbf{M}(\theta) = \overline{d\mathbf{M}(-\theta)}$. Introducing the (matrix) function $\mathbf{N}(\theta) = \mathbf{M}(\theta+) - \mathbf{M}(-\theta-)$ (which is real and symmetric) the regression spectrum S is defined as the set of points in 0 to π for which $d\mathbf{N}(\theta)$ is not the null matrix. It is then shown by Grenander & Rosenblatt that the necessary and sufficient condition that the L.S. and B.L.U. estimates of $\boldsymbol{\beta}$ should have, asymptotically, the same covariance matrix is that S should consist of $q \leqslant k$ points of increase $\theta_1, \theta_2, \ldots, \theta_q$ for which

$$\left.\begin{aligned} d\mathbf{N}(\theta_i)\,\mathbf{N}(\pi)^{-1}\,d\mathbf{N}(\theta_j) &= \text{null matrix} \quad (i \neq j) \\ &= d\mathbf{N}(\theta_i) \quad (i = j). \end{aligned}\right\} \tag{6}$$

If we now put

$$\mathbf{T}(\theta_j) = \mathbf{N}(\pi)^{-1}\,d\mathbf{N}(\theta_j),$$

we shall have

$$\sum_j \mathbf{T}(\theta_j) = \mathbf{N}(\pi)^{-1} \int_0^{\pi} d\mathbf{N}(\theta) = \mathbf{I},$$

and it follows from (6) that the $\mathbf{T}(\theta_j)$ are idempotent.

Grenander & Rosenblatt's theorem is proved under certain regularity conditions on the spectral function of the process generating the ϵ_t. In particular, the spectral density must not have a zero at any of the points θ_j. An example where the failure of this last condition makes the theorem, quoted above, invalid is given in Grenander (1951*b*, p. 568).

To apply this result to the problem of testing for serial correlation in the residuals (when on the null hypothesis the ϵ_t are independent normal variates with zero mean) let us consider the case where the test statistic is

$$r = \frac{\mathbf{z}'\mathbf{A}\mathbf{z}}{\mathbf{z}'\mathbf{z}},$$

$\mathbf{z}$ being given by (2) and $\mathbf{A} = [a_{\mu-\nu}]$ by

$$a_{\mu-\nu} = \frac{1}{2\pi}\int_{-\pi}^{\pi} g(\theta)\, e^{i(\mu-\nu)\theta}\, d\theta, \tag{7}$$

where $g(\theta)$ is a bounded (integrable) even function.

In practice the matrix $\mathbf{A}$ will be such that $\{g(\theta)\}^s$ (s a positive integer) may be expanded in a convergent Fourier series. For example, $\mathbf{A}_d$ is near to a matrix for which

$$g(\theta) = 2(1-\cos\theta).$$

Then we shall have $$a^{(s)}_{\mu-\nu} = \frac{1}{2\pi}\int_{-\pi}^{\pi} \{g(\theta)\}^s\, e^{i(\mu-\nu)\theta}\, d\theta,$$

where $a^{(s)}_{\mu-\nu}$ is the element in the μth row and νth column of the sth power of the doubly infinite matrix obtained from $\mathbf{A}$ by allowing μ and ν to run from $-\infty$ to ∞. Moreover, by a theorem due to Szego (see Grenander, 1951*a*) it follows that the latent roots of $\mathbf{A}$ are asymptotically the equidistributed ordinates of $g(\theta)$, in the sense that

$$\lim_{n\to\infty}\frac{1}{n}\{\lambda_n^s(1)+\lambda_n^s(2)+\ldots+\lambda_n^s(n)\} = \frac{1}{2\pi}\int_{-\pi}^{\pi}\{g(\theta)\}^s\, d\theta, \tag{8}$$

where the $\lambda_n(j)$ are the latent roots of $\mathbf{A}$ and lie in the range of variation of $g(\theta)$.

Consider the quantity

$$\frac{\boldsymbol{\phi}'_\mu \mathbf{A}\boldsymbol{\phi}_\nu}{\{\boldsymbol{\phi}'_\mu\boldsymbol{\phi}_\mu \cdot \boldsymbol{\phi}'_\nu\boldsymbol{\phi}_\nu\}^{\frac{1}{2}}},$$

where the $\boldsymbol{\phi}_\mu$ satisfy the conditions of Grenander & Rosenblatt's theorem. This is easily seen to converge, as n increases, to

$$\sum_{-\infty}^{\infty} a_m r_{\mu\nu}(-m).$$

We then have, from (5),

$$\begin{aligned}\lim_{n\to\infty}\frac{\boldsymbol{\phi}'_\mu \mathbf{A}\boldsymbol{\phi}_\nu}{\{\boldsymbol{\phi}'_\mu\boldsymbol{\phi}_\mu \cdot \boldsymbol{\phi}'_\nu\boldsymbol{\phi}_\nu\}^{\frac{1}{2}}} &= \sum_{-\infty}^{\infty} a_m \int_{-\pi}^{\pi} e^{-im\theta}\, dm_{\mu\nu}(\theta)\\ &= \int_{-\pi}^{\pi} g(-\theta)\, dm_{\mu\nu}(\theta)\\ &= \int_0^{\pi} g(\theta)\, dn_{\mu\nu}(\theta)\\ &= \sum_{j=1}^{q} g(\theta_j)\, n_{\mu\nu}(j),\end{aligned}$$

where $n_{\mu\nu}(j)$ is the jump in $n_{\mu\nu}(\theta) = m_{\mu\nu}(\eta) - m_{\mu\nu}(-\theta)$ at $\theta = \theta_j$.

In the same way it may be shown that we shall also have

$$\lim_{n\to\infty}\frac{\boldsymbol{\phi}'_\mu \mathbf{A}^s\boldsymbol{\phi}_\nu}{\{\boldsymbol{\phi}'_\mu\boldsymbol{\phi}_\mu \cdot \boldsymbol{\phi}'_\nu\boldsymbol{\phi}_\nu\}^{\frac{1}{2}}} = \sum_{j=1}^{q} g(\theta_j)^s\, n_{\mu\nu}(j). \tag{9}$$

As Durbin & Watson (1950) show, the moments of r may be obtained from the quantities

$$\Sigma\nu_j^s = \text{Trace}\,(\mathbf{QA}^s)$$

(see also Whittle, 1951). Here the ν_j are the latent roots of $\mathbf{QA}$ while $\mathbf{Q}$ was defined in (2).

We may without loss of generality take the vectors $\mathbf{x}_i$ in (1) to be orthonormal. We then require

$$\operatorname{Tr}[\mathbf{A} - \sum_i \mathbf{x}_i \mathbf{x}_i' \mathbf{A}]^s.$$

For ease of printing in what follows it will be convenient to alter our notation, putting $\mathbf{x}(i)$ for $\mathbf{x}_i$, $n(\mu, \nu, j)$ for $n_{\mu\nu}(j)$, $g(j)$ for $g(\theta_j)$ and $\mathbf{T}(j)$ for $\mathbf{T}(\theta_j)$.

We may expand the matrix product just written and evaluate the trace as the sum of a number of traces. Remembering that the trace of the product of two matrices is independent of the order of their multiplication we may consider the typical term in this sum as

$$\operatorname{Tr}\{\mathbf{A}^a(\sum_i \mathbf{x}(i)\,\mathbf{x}(i)'\,\mathbf{A})^b\,\mathbf{A}^c \dots (\sum_i \mathbf{x}(i)\,\mathbf{x}(i)'\,\mathbf{A})^f\},$$

where $a+b+c+\dots+f = s$. Putting the sum of the powers of the terms $\Sigma \mathbf{x}(i)\,\mathbf{x}(i)'\,\mathbf{A}$ equal to $t \leqslant s$ we may write this as

$$\Sigma\{\mathbf{x}(i_t)'\,\mathbf{A}^{a+1}\mathbf{x}(i_1)\}\{\mathbf{x}(i_1)'\,\mathbf{A}\mathbf{x}(i_2)\} \dots \{\mathbf{x}(i_b)'\,\mathbf{A}^{c+1}\mathbf{x}(i_{b+1})\} \dots \{\mathbf{x}(i_{t-1})'\,\mathbf{A}\mathbf{x}(i_t)\},$$

where the summation is over all $i_1, i_2, \dots, i_t$ running from 1 to k. In the limit, if the $\mathbf{x}(i)$ satisfy the conditions of Grenander & Rosenblatt's theorem, we have from (9)

$$\Sigma \left\{\sum_{j=1}^{q} g(j)^{a+1} n(i_t, i_1; j)\right\} \left\{\sum_{j=1}^{q} g(j)\, n(i_1, i_2; j)\right\} \dots \left\{\sum_{j=1}^{q} g(j)^{c+1} n(i_b, i_{b+1}; j)\right\} \dots \left\{\sum_{j=1}^{q} g(j)\, n(i_{t-1}, i_t; j)\right\}.$$

Changing the order of summation we get

$$\Sigma\{g(j_1)^{a+1} g(j_2) \dots g(j_{b+1})^{c+1} \dots g(j_t)\, \Sigma n(i_t, i_1; j_1)\, n(i_1, i_2; j_2) \dots n(i_b, i_{b+1}; j_{b+1}) \dots n(i_{t-1}, i_t; j_t)\},$$

where the outer summation is over $j_1, j_2, \dots, j_t$ running from 1 to q and the inner one is over $i_1, i_2, \dots, i_t$ running from 1 to k. But the inner summation is seen to be

$$\operatorname{Tr} \mathbf{T}(j_1)\,\mathbf{T}(j_2) \dots \mathbf{T}(j_t),$$

since

$$\begin{aligned} \mathbf{N}(\pi) = \mathbf{R}(0) &= \mathbf{I} \\ &= 0 \qquad \text{unless} \quad j_1 = j_2 = \dots = j_t, \\ &= \operatorname{Tr} \mathbf{T}(j) \quad \text{if} \qquad j_1 = j_2 = \dots = j_t = j \end{aligned}$$

(from (6) and the fact that the $\mathbf{T}(j)$ are idempotent). Hence

$$\lim_{n\to\infty} \operatorname{Tr}\{\mathbf{A}^a(\sum_i \mathbf{x}(i)\,\mathbf{x}(i)'\,\mathbf{A})^b\,\mathbf{A}^c \dots (\sum_i \mathbf{x}(i)\,\mathbf{x}(i)'\,\mathbf{A})^f\} = \sum_{j=1}^{q} g(j)^s \operatorname{Tr} \mathbf{T}(j),$$

and is independent of the partition of s into integers $a, b, c, \dots, f$ provided $t > 0$.

Reverting to our original notation we then have

$$\begin{aligned} &\lim_{n\to\infty} \operatorname{Tr}[\mathbf{A} - \sum_i \mathbf{x}_i \mathbf{x}_i' \mathbf{A}]^s \\ &\quad = \operatorname{Tr} \mathbf{A}^s + \sum_{r=1}^{s} (-)^r \binom{s}{r} \left\{\sum_{j=1}^{q} g(\theta_j)^s \operatorname{Tr} \mathbf{T}(\theta_j)\right\} \\ &\quad = \operatorname{Tr} \mathbf{A}^s - \sum_{j=1}^{q} g(\theta_j)^s \operatorname{Tr} \mathbf{T}(\theta_j). \end{aligned} \tag{10}$$

It is easy to see that $\operatorname{Tr} \mathbf{T}(\theta_j)$ is invariant under the orthonormalization so that this formula is perfectly general. In fact under the orthonormalization the $\mathbf{T}(\theta_j) = \mathbf{N}(\pi)^{-1} d\mathbf{N}(\theta_j)$ will be replaced by $\mathbf{P}' d\mathbf{N}(\theta_j)\,\mathbf{P}$, where

$$\mathbf{P}'\mathbf{N}(\pi)\,\mathbf{P} = \mathbf{I}.$$

It can be easily seen that the $\mathbf{P}' d\mathbf{N}(\theta_j)\,\mathbf{P}$ annihilate each other and are idempotent. They are also symmetric and non-negative definite, however, so that each must have units in certain places in the principal diagonal and zeros elsewhere, no two having unity in the same place. (This fact could be used to obtain an alternative proof of (10).) It follows therefore that

$$\operatorname{Tr} \mathbf{P}' d\mathbf{N}(\theta_j)\,\mathbf{P} = \operatorname{Tr} \mathbf{N}(\pi)^{-1} d\mathbf{N}(\theta_j) = \operatorname{Tr} \mathbf{T}(\theta_j) = p_j \quad (p_j \text{ integral}),$$

where

$$\sum_1^q p_j = k.$$

We then have, under the conditions of Grenander & Rosenblatt's theorem,

$$\lim_{n\to\infty} \{n^{-1} \operatorname{Tr} (\mathbf{QA})^s\} = n^{-1} \left\{ \operatorname{Tr} \mathbf{A}^s - \sum_1^q p_j g(\theta_j)^s \right\}$$

$$= \frac{1}{2\pi} \left\{ \int_{-\pi}^{\pi} g(\theta)^s \, d\theta - \sum_1^q p_j g(\theta_j)^s \frac{2\pi}{n} \right\}$$

from (8), and it can be seen that, asymptotically, the distribution of r is the same as that which would obtain if the regression vectors had been latent vectors corresponding to latent roots $g(\theta_j)$, repeated p_j times. As n increases, of course, $g(\theta_j)$ will be arbitrarily near to p_j latent roots of A.

For example, if

$$x_{2j-1}(t) = \left\{ t^p \sin \frac{2\pi g_j t}{n} \right\}, \quad x_{2j}(t) = \left\{ t^p \cos \frac{2\pi g_j t}{n} \right\} \quad (j = 1, \ldots, q),$$

then $\mathbf{R}_h$ has $\cos\left(h \dfrac{2\pi g_j}{n}\right)$, repeated twice, in the principal diagonal and zeros elsewhere so that $d\mathbf{T}(\theta)$ is null except at $\theta = 2\pi g_j/n$, where it has two units in the principal diagonal and zeros elsewhere. If as $\mathbf{A}$ we choose $\mathbf{A}_c$, the matrix occurring in the definition of the circular serial correlation coefficient (see Anderson & Anderson, 1950), we obtain

$$\lim_{n\to\infty} \left\{ \frac{1}{n} \operatorname{Tr} (\mathbf{QA}_c)^s \right\} = \frac{1}{n} \left\{ \operatorname{Tr} \mathbf{A}_c^s - \sum_1^q 2 \left(\cos \frac{2\pi g_j}{n} \right)^s \right\},$$

since $g(\theta)$ now becomes $\cos\theta$. For suitable g_j the statistic $r_c = \dfrac{\mathbf{z}'\mathbf{A}_c\mathbf{z}}{\mathbf{z}'\mathbf{z}}$ could then be referred to the tables prepared by Anderson & Anderson. The validity of the approximation for small n would need to be checked, of course. The example is not very realistic however, and will not be considered further here. A more real case would be that where p is allowed to run through the integers from 0 to k (say). Now, however, we shall have

$$\lim_{n\to\infty} \left\{ \frac{1}{n} \operatorname{Tr} (\mathbf{QA}_c)^s \right\} = \frac{1}{n} \left\{ \operatorname{Tr} \mathbf{A}_c^s - (k+1) \sum_1^q 2 \left(\cos \frac{2\pi g_j}{n} \right)^s \right\},$$

and this does not correspond to any statistic whose significance points are at present tabulated.

3. REGRESSION ON THE ORTHOGONAL POLYNOMIALS

Here we shall use the matrix $\mathbf{A}_d$. Though $\mathbf{A}_d$ does not satisfy (7) it is very near to a matrix which does and the difference will not matter asymptotically. We now have, in the limit, $g(\theta) = 2(1-\cos\theta)$. For the orthogonal polynomials there is only one point of increase in $\mathbf{N}(\theta)$, at the origin, for which $\mathbf{T}(0)$ is the unit matrix so that

$$\operatorname{Tr}[\mathbf{Q}\mathbf{A}_d]^s \to \operatorname{Tr}\mathbf{A}_d^s - kg(0) = \operatorname{Tr}\mathbf{A}_d^s.$$

Since d is necessarily less than d_u (see (3)) and the roots $\lambda_1, \lambda_2, \ldots, \lambda_{k-1}$ all tend to zero as $n \to \infty$, it appears that the effect of the regression is to eliminate the k smallest roots from the spectrum of $\mathbf{A}_d$ and the appropriate significance point is, asymptotically, that given by the upper bound to the significance point as tabulated by Durbin & Watson (1951). (Note that in their tables they put $k' = k-1$.)

To check on the adequacy of the approximation for n small the mean μ and variance σ^2 of d and d_u have been calculated for certain n and k' $(= k-1)$ and are shown in Table 1.

Table 1

u \ k'		1		3		5	
		μ	σ^2	μ	σ^2	μ	σ^2
15	d	2·150	0·225	2·474	0·194	2·798	0·145
	d_u	2·150	0·224	2·491	0·190	2·860	0·137
20	d	2·110	0·177	2·345	0·164	2·578	0·147
	d_u	2·110	0·177	2·354	0·162	2·620	0·137
25	d	2·086	0·146	2·271	0·139	2·456	0·130
	d_u	2·086	0·146	2·275	0·138	2·481	0·124
30	d	2·071	0·124	2·222	0·120	2·376	0·114
	d_u	2·071	0·124	2·224	0·119	2·392	0·111

It appears that the true significance point will not differ very greatly from the significance point for d_u. Even in the case $n = 15$, $k' = 5$ the deviation will be only about 0·06. (Here the 5 % bounds are 0·56 and 2·21.) Having regard to the certainty of deviations from normality and the fact that critical decisions will rarely be made when the observed value falls very near to the significance point, it seems that the upper bound to the significance point as tabulated by Durbin & Watson may nearly always be used as an adequate approximation to the true significance point. When the observed value is very near to this upper bound the significance point could always be located more accurately by using $\frac{1}{4}d$ as a Beta variate, as suggested by Durbin & Watson.

The work of Durbin & Watson showed, of course, that the bounds to the significance point differed from the true significance point by quantities of the order n^{-1}. What we have here done is to show that in the case of the orthogonal polynomials the upper bounds differ from the true significance points by quantities of the order n^{-2}.

4. Mixed regression on orthogonal polynomials and random variates

A situation which is likely to arise in practice is that in which a regression is to be computed and all the variates contain a trend component. A standard procedure is then to include the orthogonal polynomials in the regression up to the order needed to eliminate the most complicated trend. This will often result in a regression on a large number of regressors, so that the bounds test for the serial correlation of the residuals will be inconclusive. For an example see Quenouille (1952, pp. 182–6).

Consider the case of a regression on k vectors $\boldsymbol{\phi}_j$ satisfying the conditions of Grenander's theorem and p vectors $\mathbf{x}_i$ which are unrestricted. We may, without loss of generality, take the whole set $\boldsymbol{\phi}_j$, $\mathbf{x}_i$ as orthonormal. Then if the non-zero latent roots of

$$\left[\mathbf{I}-\sum_1^p \mathbf{x}_i\mathbf{x}_i'-\sum_1^k \boldsymbol{\phi}_j\boldsymbol{\phi}_j'\right]\mathbf{A}=\left[\mathbf{I}-\sum_1^p \mathbf{x}_i\mathbf{x}_i'\right]\left[\mathbf{I}-\sum_1^k \boldsymbol{\phi}_j\boldsymbol{\phi}_j'\right]\mathbf{A}$$

are

$$\mu_1, \mu_2, \ldots, \mu_{n-k-p},$$

while those of $\left[\mathbf{I}-\sum_1^k \boldsymbol{\phi}_j\boldsymbol{\phi}_j'\right]\mathbf{A}$ are

$$\nu_1, \nu_2, \ldots, \nu_{n-k},$$

Durbin & Watson (1950) show that, when each set is numbered so that its members are increasing,

$$\nu_i \leqslant \mu_i \leqslant \nu_{i+p}.$$

Thus a lower bound to $r=\dfrac{\mathbf{z}'\mathbf{A}\mathbf{z}}{\mathbf{z}'\mathbf{z}}$ is

$$r_l=\frac{\sum_1^{n-k-p}\nu_i\zeta_i^2}{\sum_1^{n-k-p}\zeta_i^2}. \tag{11}$$

We have from (10)

$$\lim_{n\to\infty}\sum_{i=1}^{n-k-p}\nu_i^s=\operatorname{Tr}\mathbf{A}^s-\sum_1^q\{g(\theta_j)\}^s\operatorname{Tr}\mathbf{T}(\theta_j)-\sum_{j=0}^{p-1}\nu_{n-k-j}^s$$

$$=\sum_1^n\lambda_j^s-\sum_1^q\{g(\theta_j)\}^s\operatorname{Tr}\mathbf{T}(\theta_j)-\sum_{j=0}^{p-1}\nu_{n-k-j}^s.$$

But from the fact that

$$\lambda_i\leqslant\nu_i\leqslant\lambda_{i+k},$$

it is clear that $\sum_{j=0}^{p-1}\nu_{n-k-j}$ will differ from $\sum_0^{p-1}\lambda_{n-j}$ by a quantity of order n^{-1} so that to our order of approximation to the moments of r_l we may put

$$\sum_1^{n-k-p}\nu_i^s=\sum_1^{n-p}\lambda_j^s-\sum_1^q\{g(\theta_j)\}^s\operatorname{Tr}\mathbf{T}(\theta_j).$$

In the case where $\operatorname{Tr}\mathbf{T}(\theta_j)$ is the appropriate multiplicity of the root corresponding to $g(\theta_j)$ a lower bound to the significance point for r may therefore be obtained from

$$r_l'=\frac{\Sigma'\lambda_j\zeta_j^2}{\Sigma'\zeta_j^2},$$

where in Σ' the terms corresponding to the latent roots other than the p greatest and those corresponding to the $g(\theta_j)$ appear.

For the case of the orthogonal polynomials (using $\mathbf{A}_d$) this bound becomes

$$d_l' = \frac{\sum_{1}^{n-k-p} \lambda_{j+k-1}\zeta_j^2}{\sum_{1}^{n-k-p} \zeta_j^2} \tag{12}$$

(the latent root, zero, corresponding to the constant term having been omitted, to accord with Durbin & Watson's notation, as in (3)).

This is not a true lower bound, for finite n since (for example) $\sum_{1}^{n-k} \nu_i < \sum_{1}^{n-k} \lambda_{i+k-1}$. However, the considerations of the previous section indicate that the error involved in using it will be very small. Moreover, one can only be wrong in using it as a lower bound when the p vectors $\mathbf{x}_i$ are near to the latent vectors corresponding to the largest latent roots of $\mathbf{A}_d$. This implies, for example, high negative serial correlation and would probably be recognized before the analysis was commenced. However, this discussion is largely academic, since the significance points for d_l' in (12) are not tabulated. Until such time as the necessary work can be done an expedient (which will narrow the bounds, though not as much as d_l' will narrow them) is to use

$$d_l'' = \frac{\sum_{1}^{n-p-1} \lambda_i\zeta_i^2}{\sum_{1}^{n-p-1} \zeta_i^2}.$$

This is obtained from (12) by adding back the terms corresponding to the $(k-1)$ smallest latent roots (the kth vector in the set being the vector corresponding to the constant term). For the range of values of n, p and k occurring in practice this will provide a lower bound for d_l'.* The significance points for d_l'' are tabulated by Durbin & Watson (see (3)) and can be obtained from their tables for d_l putting $p = k'$.

We shall close this section with an example which originally suggested the problem. In Quenouille (1952, p. 183) an example is given of the regression of U.S. fertilizer consumption on an index of farm income over the years 1911–47. The two series contain a decided trend which appears to need a fifth degree polynomial for its removal. In this case, therefore, $p = 1$, $k - 1 = k' = 5$ and $n = 37$. In Table 2 the 5 % points for the four statistics d_l, d_l'', d_l' and d_u are shown for these p, k and n.

Table 2. 5 % *points*

d_l	d_l''	d_l'	d_u
1·14	1·42	1·75	1·88

These significance points were obtained by using a Beta distribution with the appropriate mean and variance but will certainly not differ from the exact values by more than 0·02.

Quenouille quotes the first serial correlation of the residuals as 0·229, which corresponds to a value of d slightly less than 1·54. Using Durbin & Watson's original bounds no conclusion may be reached. Using the conservative lower bound d_l'' (obtained from Durbin & Watson's tables) this is still so. However, using the bound given by d_l' it can unquestionably

* For $kn^{-1} \leqslant \frac{1}{3}$ this will certainly be so.

be said that the serial correlation is significant at the 5 % point (as Quenouille conjectured). In this case the random regressor is positively serially correlated and there can be no doubt of the validity of the result based on d_l'.

I should like to thank Dr G. S. Watson for some helpful discussion in connexion with this work.

REFERENCES

ANDERSON, T. W. (1948). *Skand. AktuarTidksr.* **31**, 88.
ANDERSON, R. L. & T. W. (1950). *Ann. Math. Statist.* **21**, 59.
DURBIN, J. & WATSON, G. S. (1950). *Biometrika*, **37**, 409.
DURBIN, J. & WATSON, G. S. (1951). *Biometrika*, **38**, 159.
GRENANDER, U. (1951*a*). *Ark. Mat.* **1**, 503.
GRENANDER, U. (1951*b*). *Ark. Mat.* **1**, 555.
GRENANDER, U. (1954). *Ann. Math. Statist.* **25**, 252.
GRENANDER, U. & ROSENBLATT, M. (1954). *Proc. Nat. Acad. Sci., Wash.*, **40**, 812.
HANNAN, E. J. (1955). *Biometrika*, **42**, 133.
QUENOUILLE, M. H. (1952). *Associated Measurements*. London: Butterworth.
WHITTLE, P. (1951). *Hypothesis Testing in Time Series Analysis*. Uppsala: Almqvist and Wiksells.

ON THE ANALYSIS OF MULTIPLE REGRESSION IN k CATEGORIES

By S. KULLBACK and H. M. ROSENBLATT*

The George Washington University and Bureau of Ordnance, U.S. Navy Dept.†

SUMMARY. The general model, an information theoretic approach and solution to problems of test of hypotheses concerning sets of partial regression coefficients from k categories each involving $p+1$ variates, is presented and applied to certain data. The significance test is the analysis of variance ratio. It is shown that Carter's (1949) problem involving a 'correlation effect' among the ith members of each category reduces to a special case. The case of stochastic dependence among categories has been included by Kullback (1956) in his discussion of the multivariate linear hypothesis.

1. Introduction

We first consider the problem of the normal linear hypothesis (Kolodziejczyk, 1935) from the viewpoint of information theory, and then apply the results and procedure to the problem of tests of hypotheses about sets of partial regression coefficients. It is believed that this approach has some merit, at least from the pedagogical point of view, and has served as the basis for teaching some aspects of Design of Experiments at the George Washington University. Although the general theory given has wider applicability, the specific results given here may serve some needs for k category multiple regression analysis. Aside from Carter's (1949) paper, the k category multiple regression case does not seem to be covered in the literature, and has not been found in recent texts on statistical theory or method (cf. Kempthorne, 1952; Kendall, 1946; Pearson & Wilks, 1933; Rao, 1952; Welch, 1935; Williams, 1953). Study of the problem by Rosenblatt (1953) was stimulated by the needs of an applied problem at the Naval Ordnance Laboratory, White Oak, Md. Complete results and a general theory are presented here. Matrix notation and theory are used and the results are illustrated by application to certain numerical data. Matrices will be denoted by upper case bold face type, e.g. $\mathbf{A} = (a_{ij})$, $\mathbf{X}_1 = (x_{1ij})$, etc. $(i = 1, 2, \ldots, m; j = 1, 2, \ldots, n)$, the vectors or 1 row or column matrices will be denoted by lower case bold face type, e.g. $\mathbf{x}' = (x_1, x_2, \ldots, x_k)$, $\boldsymbol{\mu}_1' = (\mu_{11}, \mu_{12}, \ldots, \mu_{1p})$, etc.

1·1. *Information theory approach.* Let $f_1(x)$ and $f_2(x)$ be the probability densities of populations specified by the statistical hypotheses H_1 and H_2, respectively. The mean information per observation from the population with probability density $f_1(x)$ for discrimination for H_1 against H_2, is defined by‡

$$I(1:2) = \int f_1(x) \log \frac{f_1(x)}{f_2(x)}\, dx, \tag{1·1}$$

where x can be taken as a vector variable (multivariate). The divergence between H_1 and H_2, a measure of the difficulty of discriminating between them, is defined by‡

$$J(1,2) = \int [f_1(x) - f_2(x)] \log \frac{f_1(x)}{f_2(x)}\, dx. \tag{1·2}$$

* Now with the Office of Naval Research, U.S. Navy Department, Washington, D.C.

† Presented 24 March 1955 at a Bureau of Ordnance Seminar of Statisticians held at the University of Chicago.

‡ See Kullback & Leibler (1951) for the case of general populations.

For the special case when $f_1(x)$ and $f_2(x)$ are p-variate normal densities, with the same matrix of variances and covariances, $\mathbf{\Sigma}$, and mean values

$$E_1(x) = \boldsymbol{\mu}_1' = (\mu_{11}, \mu_{12}, \ldots, \mu_{1p}) \quad \text{and} \quad E_2(x) = \boldsymbol{\mu}_2' = (\mu_{21}, \mu_{22}, \ldots, \mu_{2p}),$$

it is found (Kullback, 1952) that (1·1) and (1·2) yield

$$2I(1:2) = J(1,2) = (\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2)' \, \mathbf{\Sigma}^{-1} (\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2), \tag{1·3}$$

where the right-hand side of (1·3) is proportional to Mahalanobis's generalized distance (see, for example, Rao, 1952).

1·2. *The linear hypothesis—single category.* We now consider the set of n observations on $p+1$ variates

$$\mathbf{z} = \mathbf{y} - \mathbf{X}\boldsymbol{\beta}, \tag{1·4}$$

where

$$\mathbf{z}' = (z_1, z_2, \ldots, z_n), \quad \mathbf{y}' = (y_1, y_2, \ldots, y_n),$$

$$\boldsymbol{\beta}' = (\beta_1, \beta_2, \ldots, \beta_p), \quad \mathbf{X} = (x_{ir}) \quad (i = 1, 2, \ldots, n;\ r = 1, 2, \ldots, p;\ p < n),$$

such that

(*a*) the z's are independent, normally distributed variables with zero means and variance σ^2,

(*b*) the x_{ir}'s are considered to be known,

(*c*) $\mathbf{X}$ is of rank p,

(*d*) $\boldsymbol{\beta} = \boldsymbol{\beta}^1$ and $\boldsymbol{\beta} = \boldsymbol{\beta}^2$ are parameter matrices whose values are respectively given by the hypotheses H_1 and H_2,

(*e*) $E_1(\mathbf{y}) = \mathbf{X}\boldsymbol{\beta}^1$ and $E_2(\mathbf{y}) = \mathbf{X}\boldsymbol{\beta}^2$.

It is found that (1·3) yields for this case,

$$\begin{aligned} J(1,2) &= (\mathbf{X}\boldsymbol{\beta}^1 - \mathbf{X}\boldsymbol{\beta}^2)' \, (\sigma^2\mathbf{I})^{-1} (\mathbf{X}\boldsymbol{\beta}^1 - \mathbf{X}\boldsymbol{\beta}^2) \\ &= (\boldsymbol{\beta}^1 - \boldsymbol{\beta}^2)' \, \mathbf{S} (\boldsymbol{\beta}^1 - \boldsymbol{\beta}^2)/\sigma^2, \end{aligned} \tag{1·5}$$

where $\mathbf{S} = \mathbf{X}'\mathbf{X}$ is a $p \times p$ matrix of rank p and $\mathbf{I}$ is the $n \times n$ identity matrix.

Suppose H_1 imposes no restriction on $\boldsymbol{\beta}$ and H_2 provides some specific hypothetical value $\boldsymbol{\beta} = \boldsymbol{\beta}^2$. We estimate $J(1,2)$ by replacing the parameters by the best unbiased estimates appropriate to the hypotheses. That this procedure is based on the principle of discriminating between a null hypothesis and the alternative hypothesis, by using that distribution corresponding to the alternative hypothesis which for the sample values provides the least information for discrimination against the null hypothesis—is shown and discussed more generally by one of us (Kullback, 1956). The classical least-squares procedure of minimizing

$$\mathbf{z}'\mathbf{z} = (\mathbf{y}' - \boldsymbol{\beta}'\mathbf{X}') \, (\mathbf{y} - \mathbf{X}\boldsymbol{\beta})$$

leads to the normal equations

$$\mathbf{S}\hat{\boldsymbol{\beta}} = \mathbf{X}'\mathbf{y} \tag{1·6}$$

whose solutions are minimum variance unbiased estimates (cf. Kempthorne, 1952; Kolodziejczyk, 1935; Plackett, 1949), so that

$$\hat{J}(1,2) = (\hat{\boldsymbol{\beta}}^1 - \hat{\boldsymbol{\beta}}^2)' \, \mathbf{S} (\hat{\boldsymbol{\beta}}^1 - \hat{\boldsymbol{\beta}}^2)/\hat{\sigma}^2, \tag{1·7}$$

where

$$(n-p)\,\hat{\sigma}^2 = \hat{\mathbf{z}}'\hat{\mathbf{z}} = (\mathbf{y}' - \hat{\boldsymbol{\beta}}'^1\mathbf{X}') \, (\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}}^1) = \mathbf{y}'\mathbf{y} - \hat{\boldsymbol{\beta}}'^1\mathbf{S}\boldsymbol{\beta}^1.$$

In particular, for the common null hypothesis, H_2: $\boldsymbol{\beta} = \boldsymbol{\beta}^2 = \mathbf{0}$, (1·7) becomes

$$\hat{J}(1,2) = \hat{\boldsymbol{\beta}}'^1\mathbf{S}\hat{\boldsymbol{\beta}}^1/\hat{\sigma}^2. \tag{1·8}$$

It is a known result in regression theory that the components of $\hat{\boldsymbol{\beta}}^1$ as linear functions of the z's are normally distributed with covariance matrix given by $\sigma^2\mathbf{S}^{-1}$ (cf. Kempthorne, 1952).

Under the null hypothesis, that $\boldsymbol{\beta} = \boldsymbol{\beta}^2 = \mathbf{0}$, $\hat{J}(1,2)$ in (1·8) is therefore Hotelling's Generalized Student Ratio (Wilks, 1943, p. 238), or

$$\hat{J}(1,2) = pF, \tag{1·9}$$

where F has the analysis of variance distribution with $n_1 = p$ and $n_2 = n-p$ degrees of freedom. These results may also be summarized in the usual analysis of variance table as given in Table 1·1, where

$$\hat{\boldsymbol{\beta}}'^1\mathbf{S}\hat{\boldsymbol{\beta}}^1 = \hat{\boldsymbol{\beta}}'^1\mathbf{X}'\mathbf{y} = \mathbf{y}'\mathbf{X}\mathbf{S}^{-1}\mathbf{X}'\mathbf{y}.$$

For the more general case, corresponding to the hypothesis, H_2: $\boldsymbol{\beta} = \boldsymbol{\beta}^2 \neq \mathbf{0}$, (1·9) still holds with $\hat{J}(1,2)$ given by (1·7).

Table 1·1

Variation due to	D.F.	Sum of squares
Linear regression	p	$\hat{\boldsymbol{\beta}}'^1\mathbf{S}\hat{\boldsymbol{\beta}}^1 = \hat{\sigma}^2\hat{J}(1,2)$
Difference	$n-p$	$\mathbf{y}'\mathbf{y} - \hat{\boldsymbol{\beta}}'^1\mathbf{S}\hat{\boldsymbol{\beta}}^1 = (n-p)\,\hat{\sigma}^2$
Total	n	$\mathbf{y}'\mathbf{y}$

2. SUB-HYPOTHESES, SINGLE CATEGORY*

Suppose we partition the parameters $\boldsymbol{\beta}$ into two groups, which we will denote as $\boldsymbol{\beta}_1$ and $\boldsymbol{\beta}_2$, so that in place of (1·4) we now consider

$$\mathbf{z} = \mathbf{y} - (\mathbf{X}_1, \mathbf{X}_2)\begin{pmatrix}\boldsymbol{\beta}_1\\ \boldsymbol{\beta}_2\end{pmatrix}, \tag{2·1}$$

where $\boldsymbol{\beta} = \begin{pmatrix}\boldsymbol{\beta}_1\\ \boldsymbol{\beta}_2\end{pmatrix}$ and $\mathbf{X} = (\mathbf{X}_1, \mathbf{X}_2) = \left(\begin{array}{ccc|ccc} x_{11}, & \ldots, & x_{1q} & x_{1q+1}, & \ldots, & x_{1p} \\ \vdots & & & & & \vdots \\ x_{n1}, & \ldots, & x_{nq} & x_{nq+1}, & \ldots, & x_{np} \end{array}\right)$,

with $\mathbf{X}_1$ and $\mathbf{X}_2$ respectively of ranks q and $p-q$. The z's are still assumed to be independent and normally distributed with means zero and common variance σ^2, and corresponding to H_1 and H_2,

$$\left.\begin{aligned} E_1(\mathbf{y}) &= \mathbf{X}_1\boldsymbol{\beta}_1{}^1 + \mathbf{X}_2\boldsymbol{\beta}_2{}^1, \\ E_2(\mathbf{y}) &= \mathbf{X}_1\boldsymbol{\beta}_1{}^2 + \mathbf{X}_2\boldsymbol{\beta}_2{}^2. \end{aligned}\right\} \tag{2·2}$$

It also follows that

$$\mathbf{S} = \mathbf{X}'\mathbf{X} = \begin{pmatrix}\mathbf{S}_{11} & \mathbf{S}_{12}\\ \mathbf{S}_{21} & \mathbf{S}_{22}\end{pmatrix}, \tag{2·3}$$

where

$$\mathbf{S}_{11} = \mathbf{X}_1'\mathbf{X}_1, \quad \mathbf{S}_{12} = \mathbf{X}_1'\mathbf{X}_2 = \mathbf{S}_{21}', \quad \mathbf{S}_{22} = \mathbf{X}_2'\mathbf{X}_2.$$

* For an alternative treatment see Grundy (1951). We summarize the results here as a preliminary aid for the discussion in § 4.

For this case, (1·5) becomes

$$J(1,2) = (\boldsymbol{\beta}_1'^1 - \boldsymbol{\beta}_1'^2,\ \boldsymbol{\beta}_2'^1 - \boldsymbol{\beta}_2'^2)\begin{pmatrix}\mathbf{S}_{11} & \mathbf{S}_{12}\\ \mathbf{S}_{21} & \mathbf{S}_{22}\end{pmatrix}\begin{pmatrix}\boldsymbol{\beta}_1^1 - \boldsymbol{\beta}_1^2\\ \boldsymbol{\beta}_2^1 - \boldsymbol{\beta}_2^2\end{pmatrix}\Big/\sigma^2. \tag{2·4}$$

The normal equations of (1·6) under H_1 become

$$\begin{pmatrix}\mathbf{S}_{11} & \mathbf{S}_{12}\\ \mathbf{S}_{21} & \mathbf{S}_{22}\end{pmatrix}\begin{pmatrix}\hat{\boldsymbol{\beta}}_1^1\\ \hat{\boldsymbol{\beta}}_2^1\end{pmatrix} = \begin{pmatrix}\mathbf{X}_1'\\ \mathbf{X}_2'\end{pmatrix}\mathbf{y}, \tag{2·5}$$

or

$$\left.\begin{aligned}\mathbf{S}_{11}\hat{\boldsymbol{\beta}}_1^1 + \mathbf{S}_{12}\hat{\boldsymbol{\beta}}_2^1 &= \mathbf{X}_1'\mathbf{y},\\ \mathbf{S}_{21}\hat{\boldsymbol{\beta}}_1^1 + \mathbf{S}_{22}\hat{\boldsymbol{\beta}}_2^1 &= \mathbf{X}_2'\mathbf{y},\end{aligned}\right\} \tag{2·6}$$

and

$$(n-p)\,\hat{\sigma}^2 = \mathbf{y}'\mathbf{y} - (\hat{\boldsymbol{\beta}}_1'^1, \hat{\boldsymbol{\beta}}_2'^1)\begin{pmatrix}\mathbf{S}_{11} & \mathbf{S}_{12}\\ \mathbf{S}_{21} & \mathbf{S}_{22}\end{pmatrix}\begin{pmatrix}\hat{\boldsymbol{\beta}}_1^1\\ \hat{\boldsymbol{\beta}}_2^1\end{pmatrix}.$$

Let

$$\mathbf{S}_{22.1} = \mathbf{S}_{22} - \mathbf{S}_{21}\mathbf{S}_{11}^{-1}\mathbf{S}_{12}, \quad \mathbf{X}_{2.1}' = \mathbf{X}_2' - \mathbf{S}_{21}\mathbf{S}_{11}^{-1}\mathbf{X}_1',$$

then (2·6) yields

$$\hat{\boldsymbol{\beta}}_2^1 = \mathbf{S}_{22.1}^{-1}\mathbf{X}_{2.1}'\mathbf{y}, \tag{2·7}$$

$$\hat{\boldsymbol{\beta}}_1^1 = \mathbf{S}_{11}^{-1}\mathbf{X}_1'\mathbf{y} - \mathbf{S}_{11}^{-1}\mathbf{S}_{12}\hat{\boldsymbol{\beta}}_2^1. \tag{2·8}$$

It is useful to note (see, for example, Frazer, Duncan & Collar, 1938, para. 4·9) that

$$\mathbf{S}^{-1} = \begin{pmatrix}\mathbf{S}_{11} & \mathbf{S}_{12}\\ \mathbf{S}_{21} & \mathbf{S}_{22}\end{pmatrix}^{-1} = \begin{pmatrix}\mathbf{S}_{11.2}^{-1} & \mathbf{M}\\ \mathbf{M}' & \mathbf{S}_{22.1}^{-1}\end{pmatrix},$$

where the $q \times (p-q)$ matrix

$$\mathbf{M} = -\mathbf{S}_{11}^{-1}\mathbf{S}_{12}\mathbf{S}_{22.1}^{-1} = -\mathbf{S}_{11.2}^{-1}\mathbf{S}_{12}\mathbf{S}_{22}^{-1},$$

so that in the applications, the elements of the matrix $\mathbf{S}_{11.2}^{-1}$ or $\mathbf{S}_{22.1}^{-1}$ are already obtained once the matrix $\mathbf{S}^{-1}$ is obtained.

Suppose now that in particular we want to test the null hypothesis H_2: $\boldsymbol{\beta} = \boldsymbol{\beta}^2 = \begin{pmatrix}\boldsymbol{\beta}_1^2\\ \mathbf{0}\end{pmatrix}$, that is, $\boldsymbol{\beta}_2^2 = \mathbf{0}$, with no restrictions on $\boldsymbol{\beta}_1^2$, while under the alternative hypothesis H_1: $\boldsymbol{\beta} = \boldsymbol{\beta}^1 = \begin{pmatrix}\boldsymbol{\beta}_1^1\\ \boldsymbol{\beta}_2^1\end{pmatrix}$ with no restrictions on the parameters. Again we estimate $J(1,2)$ by replacing the parameters by the best unbiased estimates appropriate to the hypotheses. Under H_1 we have the previous results for $\hat{\boldsymbol{\beta}}_1^1$, $\hat{\boldsymbol{\beta}}_2^1$ and $\hat{\sigma}^2$. Under H_2 the normal equations of (1·6) now yield

$$\hat{\boldsymbol{\beta}}_1^2 = \mathbf{S}_{11}^{-1}\mathbf{X}_1'\mathbf{y}. \tag{2·9}$$

From (2·4), (2·8) and (2·9) we have that

$$\hat{\sigma}^2\hat{J}(1,2) = (-\hat{\boldsymbol{\beta}}_2'^1\mathbf{S}_{21}\mathbf{S}_{11}^{-1}, \hat{\boldsymbol{\beta}}_2'^1)\begin{pmatrix}\mathbf{S}_{11} & \mathbf{S}_{12}\\ \mathbf{S}_{21} & \mathbf{S}_{22}\end{pmatrix}\begin{pmatrix}-\mathbf{S}_{11}^{-1}\mathbf{S}_{12}\hat{\boldsymbol{\beta}}_2^1\\ \hat{\boldsymbol{\beta}}_2^1\end{pmatrix} = \hat{\boldsymbol{\beta}}_2'^1\mathbf{S}_{22.1}\hat{\boldsymbol{\beta}}_2^1. \tag{2·10}$$

It may be readily verified that

$$\mathbf{X}\mathbf{S}^{-1}\mathbf{X}' = \mathbf{X}_{2.1}\mathbf{S}_{22.1}^{-1}\mathbf{X}_{2.1}' + \mathbf{X}_1\mathbf{S}_{11}^{-1}\mathbf{X}_1', \tag{2·11}$$

that is,

$$\hat{\boldsymbol{\beta}}'^1\mathbf{S}\hat{\boldsymbol{\beta}}^1 = \hat{\boldsymbol{\beta}}_2'^1\mathbf{S}_{22.1}\hat{\boldsymbol{\beta}}_2^1 + \hat{\boldsymbol{\beta}}_1'^2\mathbf{S}_{11}\hat{\boldsymbol{\beta}}_1^2, \tag{2·11·1}$$

or

$$\hat{\boldsymbol{\beta}}'^1\mathbf{X}'\mathbf{y} = \hat{\boldsymbol{\beta}}_2'^1\mathbf{X}_{2.1}'\mathbf{y} + \hat{\boldsymbol{\beta}}_1'^2\mathbf{X}_1'\mathbf{y}. \tag{2·11·2}$$

The foregoing results may be summarized in the following analysis of variance (Table 2·1). $\hat{J}(1,2) = \hat{\boldsymbol{\beta}}_2'^1 \mathbf{S}_{22.1} \hat{\boldsymbol{\beta}}_2^1 / \hat{\sigma}^2 = (p-q)\, F$, where F has the analysis of variance distribution with $n_1 = p-q$ and $n_2 = n-p$ degrees of freedom, under the null hypothesis H_2 that $\boldsymbol{\beta}_2^2 = \mathbf{0}$.

Table 2·1

Variation due to	D.F.	Sum of squares
H_2: $\boldsymbol{\beta}'^2 = (\boldsymbol{\beta}_1'^2, \mathbf{0}')$	q	$\hat{\boldsymbol{\beta}}_1'^2 \mathbf{S}_{11} \hat{\boldsymbol{\beta}}_1^2$
Diff.	$p-q$	$\hat{\boldsymbol{\beta}}_2'^1 \mathbf{S}_{22.1} \hat{\boldsymbol{\beta}}_2^1 = \hat{\sigma}^2 \hat{J}(1,2)$
H_1: $\boldsymbol{\beta}'^1 = (\boldsymbol{\beta}_1'^1, \boldsymbol{\beta}_2'^1)$	p	$\hat{\boldsymbol{\beta}}'^1 \mathbf{S} \hat{\boldsymbol{\beta}}^1$
Diff.	$n-p$	$\mathbf{y}'\mathbf{y} - \hat{\boldsymbol{\beta}}'^1 \mathbf{S} \hat{\boldsymbol{\beta}}^1 = (n-p)\, \hat{\sigma}^2$
Total	n	$\mathbf{y}'\mathbf{y}$

3. Analysis of regression—one-way classification, k categories*

Suppose we have k categories each with n_j observations on $(y, x_1, \ldots, x_p)$ for which the general linear regression for each category is

$$z_{ji} = y_{ji} - (\beta_{j1} x_{ji1} + \ldots + \beta_{jr} x_{jir} + \ldots + \beta_{jp} x_{jip}), \tag{3·1}$$

where $j = 1, 2, \ldots, k$ categories, $i = 1, 2, \ldots, n_j$ observations for category j, $r = 1, 2, \ldots, p$ independent variables $(p < n_j)$, the z_{ji} are independent, normally distributed with zero means and common variance σ^2, and the x_{jir} are known.

The linear regressions for each category can be written as

$$\mathbf{z}_j = \mathbf{y}_j - \mathbf{X}_j \boldsymbol{\beta}_j, \tag{3·2}$$

where for $j = 1, 2, \ldots, k$

$$\mathbf{z}_j' = (z_{j1}, z_{j2}, \ldots, z_{jn_j}), \qquad \mathbf{y}_j' = (y_{j1}, y_{j2}, \ldots, y_{jn_j}),$$

$$\mathbf{X}_j = (\mathbf{x}_{j1}, \mathbf{x}_{j2}, \ldots, \mathbf{x}_{jp}), \quad \mathbf{x}_{jr}' = (x_{j1r}, x_{j2r}, \ldots, x_{jn_jr})$$

and

$$\boldsymbol{\beta}_j' = (\beta_{j1}, \beta_{j2}, \ldots, \beta_{jp}).$$

We may write the k sets of regression equations (3·2) for k categories combined as

$$\mathbf{z} = \mathbf{y} - \mathbf{X}\boldsymbol{\beta}, \tag{3·3}$$

by defining

$$\mathbf{X} = \begin{pmatrix} \mathbf{X}_1 & & & \mathbf{0} \\ & \mathbf{X}_2 & & \\ & & \ddots & \\ \mathbf{0} & & & \mathbf{X}_k \end{pmatrix}, \qquad \boldsymbol{\beta}' = (\boldsymbol{\beta}_1', \boldsymbol{\beta}_2', \ldots, \boldsymbol{\beta}_k'),$$

$$\mathbf{z}' = (\mathbf{z}_1', \mathbf{z}_2', \ldots, \mathbf{z}_k'), \qquad \mathbf{y}' = (\mathbf{y}_1', \mathbf{y}_2', \ldots, \mathbf{y}_k').$$

By the preceding definitions we consider $\boldsymbol{\beta}$ in (3·3) as a parameter matrix of all kp regression coefficients β_{jr} whether or not any of them are equal, or have a particular value including zero, according to any hypothesis.

* For the case $p = 1$, see Kendall (1946) and Welch (1935).

Suppose, however, we specify a null hypothesis with regard to certain groups or sets of the kp parameters β_{jr} among the k categories, and wish to estimate the parameters and test the hypothesis against some alternative. To distinguish between matrices or parameter vectors under various hypotheses H_α ($\alpha = 1, 2, \ldots$) we will use, where desirable for clarity or emphasis, the notation $\mathbf{X}^\alpha$, $\boldsymbol{\beta}^\alpha$ and $\mathbf{S}^\alpha = \mathbf{X}'^\alpha\mathbf{X}^\alpha$. Where this notation is not used the applicable hypothesis and definition of the matrices will be clear from the context. For any hypothesis, H_α, we will represent the linear regressions for the k categories combined under H_α as

$$\mathbf{z} = \mathbf{y} - \mathbf{X}^\alpha\boldsymbol{\beta}^\alpha, \tag{3·4}$$

where $\mathbf{z}$ and $\mathbf{y}$ are specified as in (3·3); however, we now define $\boldsymbol{\beta}^\alpha$ as the matrix of distinct regression coefficients specified by hypothesis H_α, and $\mathbf{X}^\alpha$, the matrix of x_{jir} with distinct regression effects, specified according to the regression model defined by the hypothesis for the k categories combined.

With the representation (3·4) of the k category regression under H_α, the normal equations (1·6) become

$$\mathbf{S}^\alpha\hat{\boldsymbol{\beta}}^\alpha = \mathbf{X}'^\alpha\mathbf{y} \tag{3·5}$$

and

$$\hat{\boldsymbol{\beta}}^\alpha = \mathbf{S}^{\alpha-1}\mathbf{X}'^\alpha\mathbf{y},$$

where the elements of $\mathbf{S}^\alpha = \mathbf{X}'^\alpha\mathbf{X}^\alpha$ will, of course, depend on the particular specification of the matrix $\mathbf{X}^\alpha$.

Also, equivalent to (1·7) we have, for a null hypothesis H_2 and an alternative H_1,

$$\hat{J}(1,2) = (\hat{\boldsymbol{\beta}}^1 - \hat{\boldsymbol{\beta}}^2)'\,\mathbf{S}(\hat{\boldsymbol{\beta}}^1 - \hat{\boldsymbol{\beta}}^2)/\hat{\sigma}^2 = (\hat{\boldsymbol{\beta}}'^1\mathbf{S}^1\hat{\boldsymbol{\beta}}^1 - \hat{\boldsymbol{\beta}}'^2\mathbf{S}^2\hat{\boldsymbol{\beta}}^2)/\hat{\sigma}^2, \tag{3·6}$$

where

$$\left.\begin{aligned}(N-pk)\,\hat{\sigma}^2 &= \mathbf{y}'\mathbf{y} - \hat{\boldsymbol{\beta}}'^1\mathbf{S}^1\hat{\boldsymbol{\beta}}^1,\\ N &= n_1 + n_2 + \ldots + n_k,\end{aligned}\right\} \tag{3·7}$$

and $\mathbf{S} = \mathbf{X}'\mathbf{X} = \mathbf{S}^1$ for $\mathbf{X}$ defined in (3·3). Thus, for any particular hypothesis on the sets of regression coefficients in k category regression, the estimates for the coefficients and the test of the hypothesis are readily obtained solely by proper specification of the matrices $\mathbf{X}^\alpha$ and $\boldsymbol{\beta}^\alpha$ in (3·4).

Consider the two hypotheses,

$$H_1\colon\quad \beta_{jr} = \beta_{jr} \quad (j = 1, 2, \ldots, k;\ r = 1, 2, \ldots, p), \tag{3·8}$$

i.e. the β_{jr} are different for all categories and for each $r = 1, 2, \ldots, p$; and the null hypothesis

$$H_2\colon\quad \beta_{jr} = \beta_{.r} \quad (j = 1, 2, \ldots, k;\ r = 1, 2, \ldots, p), \tag{3·9}$$

or equivalently,

$$\boldsymbol{\beta}'_j = \boldsymbol{\beta}'_. = (\beta_{.1}, \beta_{.2}, \ldots, \beta_{.p}) \quad (j = 1, 2, \ldots, k),$$

i.e. the regression coefficients are the same for the different categories for each $r = 1, 2, \ldots, p$.

Corresponding to H_1 in (3·8) the best unbiased estimate of $\boldsymbol{\beta}$ is derived from (3·5), where, for the hypothesis H_1, $\boldsymbol{\beta}^\alpha$ and $\mathbf{X}^\alpha$ in (3·4), defining the k category regression model, are the same as $\boldsymbol{\beta}$ and $\mathbf{X}$ in (3·3), or

$$\begin{pmatrix}\mathbf{S}_1 & & \mathbf{0}\\ & \ddots & \\ \mathbf{0} & & \mathbf{S}_k\end{pmatrix}\begin{pmatrix}\hat{\boldsymbol{\beta}}_1\\ \vdots\\ \hat{\boldsymbol{\beta}}_k\end{pmatrix} = \begin{pmatrix}\mathbf{X}'_1\mathbf{y}_1\\ \vdots\\ \mathbf{X}'_k\mathbf{y}_k\end{pmatrix}, \tag{3·10}$$

that is, k sets of normal equations

$$\mathbf{S}_j\hat{\boldsymbol{\beta}}_j = \mathbf{X}'_j\mathbf{y}_j \quad (j = 1, 2, \ldots, k), \tag{3·11}$$

from which

$$\hat{\boldsymbol{\beta}}_j = \mathbf{S}_j^{-1}\mathbf{X}'_j\mathbf{y}_j.$$

Corresponding to H_2 in (3·9), however, the matrices $\mathbf{X}^2$ and $\boldsymbol{\beta}^2$ of (3·4), defining the k category regression model, are

$$\mathbf{X}'^2 = (\mathbf{X}'_1, \mathbf{X}'_2, \ldots, \mathbf{X}'_k), \quad \boldsymbol{\beta}'^2 = (\beta_{.1}, \beta_{.2}, \ldots, \beta_{.p}).$$

Therefore,
$$\mathbf{S}^2 = \mathbf{X}'^2\mathbf{X}^2 = \sum_{j=1}^{k} \mathbf{X}'_j\mathbf{X}_j = \sum_{j=1}^{k} \mathbf{S}_j, \quad \mathbf{X}'^2\mathbf{y} = \sum_{j=1}^{k} \mathbf{X}'_j\mathbf{y}_j,$$

and the best unbiased estimate of $\boldsymbol{\beta}$ under H_2 is derived from (3·5) as

$$\hat{\boldsymbol{\beta}}^2 = \mathbf{S}^{2-1}\mathbf{X}'^2\mathbf{y}. \tag{3·12}$$

We also have, under H_1, corresponding to (3·7)

$$(N-pk)\,\hat{\sigma}^2 = \mathbf{y}'\mathbf{y} - \hat{\boldsymbol{\beta}}'^1\mathbf{S}^1\hat{\boldsymbol{\beta}}^1 = \sum_{j=1}^{k} (\mathbf{y}'_j\mathbf{y}_j - \boldsymbol{\beta}'_j\mathbf{S}_j\boldsymbol{\beta}_j). \tag{3·13}$$

Corresponding to (3·6) we therefore have

$$\hat{\sigma}^2\hat{J}(1,2) = \hat{\boldsymbol{\beta}}'^1\mathbf{S}^1\hat{\boldsymbol{\beta}}^1 - \hat{\boldsymbol{\beta}}'^2\mathbf{S}^2\hat{\boldsymbol{\beta}}^2 = \sum_{j=1}^{k} \hat{\boldsymbol{\beta}}'_j\mathbf{S}_j\hat{\boldsymbol{\beta}}_j - \hat{\boldsymbol{\beta}}'^2_{.}\mathbf{S}^2\hat{\boldsymbol{\beta}}^2_{.}. \tag{3·14}$$

This latter result is a direct generalization of Kendall's S_2 (1946, § 24·30).

We may therefore summarize in the analysis of variance Table 3·1. $\hat{J}(1,2) = p(k-1)\,F$, where F has the analysis of variance distribution with $p(k-1)$ and $(N-pk)$ degrees of freedom when the null hypothesis H_2 of (3·9) is true.

Table 3·1

Variance due to	D.F.	Sum of squares
H_2: $\boldsymbol{\beta}^2 = \boldsymbol{\beta}^2_{.}$	p	$\hat{\boldsymbol{\beta}}'^2\mathbf{S}^2\hat{\boldsymbol{\beta}}^2$
Diff.	$p(k-1)$	$\hat{\boldsymbol{\beta}}'^1\mathbf{S}^1\hat{\boldsymbol{\beta}}^1 - \hat{\boldsymbol{\beta}}'^2\mathbf{S}^2\hat{\boldsymbol{\beta}}^2 = \hat{\sigma}^2\hat{J}(1,2)$
H_1: $\boldsymbol{\beta}^1 = \boldsymbol{\beta}^1$	pk	$\hat{\boldsymbol{\beta}}'^1\mathbf{S}^1\hat{\boldsymbol{\beta}}^1 = \sum_{j=1}^{k} \hat{\boldsymbol{\beta}}'_j\mathbf{S}_j\hat{\boldsymbol{\beta}}_j$
Diff.	$N-pk$	$\mathbf{y}'\mathbf{y} - \hat{\boldsymbol{\beta}}'^1\mathbf{S}^1\hat{\boldsymbol{\beta}}^1 = (N-pk)\,\hat{\sigma}^2$
Total	$N = \Sigma n_j$	$\mathbf{y}'\mathbf{y}$

4. SUB-HYPOTHESIS—ONE-WAY CLASSIFICATION, k CATEGORIES

4·1. *Two-partition sub-hypothesis.* Let us partition the parameters of the matrix $\boldsymbol{\beta}_j$, for each category $j = 1, 2, \ldots, k$, into two parts

$$\boldsymbol{\beta}'_{j1} = (\beta_{j1}, \ldots, \beta_{jq}) \quad \text{and} \quad \boldsymbol{\beta}'_{j2} = (\beta_{jq+1}, \ldots, \beta_{jp})$$

of q and $p-q$ parameters respectively, $q < p$, so that $\boldsymbol{\beta}'_j = (\boldsymbol{\beta}'_{j1}, \boldsymbol{\beta}'_{j2})$.

Consider a sub-hypothesis, H_2, which states that for $j = 1, 2, \ldots, k$, the β_{jr} are different for $r = 1, 2, \ldots, q$, while for $r = q+1, q+2, \ldots, p$, there is a common value $\beta_{.r}$ for the β_{jr}; thus

$$\left.\begin{aligned} H_2: \quad \beta_{jr} &= \beta_{jr} \quad (j = 1, 2, \ldots, k;\ r = 1.\,2, \ldots .\, q), \\ \beta_{jr} &= \beta_{.r} \quad (j = 1, 2, \ldots, k;\ r = q+1, q+2, \ldots, p), \end{aligned}\right\} \tag{4·1}$$

or equivalently

$$\begin{aligned} H_2: \quad \boldsymbol{\beta}'_{j1} &= \boldsymbol{\beta}'_{j1} = (\beta_{j1}, \ldots, \beta_{jq}), \\ \boldsymbol{\beta}'_{j2} &= \boldsymbol{\beta}'_{.2} = (\beta_{.q+1}, \ldots, \beta_{.p}). \end{aligned}$$

Let H_1 remain as in (3·8), that is, the β_{jr} are different for all j and r. Then under H_1 we have the same matrix definitions and results as in § 3. However, for H_2 as given in (4·1), the matrices $\mathbf{X}^2$ and $\boldsymbol{\beta}^2$ for the k category regression model are

$$\boldsymbol{\beta}'^2 = (\boldsymbol{\beta}'_1, \boldsymbol{\beta}'_2), \quad \mathbf{X}^2 = (\mathbf{X}_1, \mathbf{X}_2),$$

where

$$\boldsymbol{\beta}'_1 = (\boldsymbol{\beta}'_{11}, \boldsymbol{\beta}'_{21}, \ldots, \boldsymbol{\beta}'_{k1}), \quad \boldsymbol{\beta}'_2 = (\beta_{.q+1}, \beta_{.q+2}, \ldots, \beta_{.p}) = \boldsymbol{\beta}'_{.2}$$

and

$$\mathbf{X}_1 = \begin{pmatrix} \mathbf{X}_{11} & & \mathbf{0} \\ & \ddots & \\ \mathbf{0} & & \mathbf{X}_{k1} \end{pmatrix}, \quad \mathbf{X}_2 = \begin{pmatrix} \mathbf{X}_{12} \\ \vdots \\ \mathbf{X}_{k2} \end{pmatrix},$$

$$\mathbf{X}_{j1} = (\mathbf{x}_{j1}, \mathbf{x}_{j2}, \ldots, \mathbf{x}_{jq}), \quad \mathbf{X}_{j2} = (\mathbf{x}_{jq+1}, \mathbf{x}_{jq+2}, \ldots, \mathbf{x}_{jp}),$$

$$\mathbf{x}'_{jr} = (x_{j1r}, x_{j2r}, \ldots, x_{jn_jr}) \quad (j = 1, 2, \ldots, k;\ r = 1, 2, \ldots, p).$$

Thus under H_2

$$\mathbf{S}^2 = \mathbf{X}'^2\mathbf{X}^2 = \begin{pmatrix} \mathbf{X}'_1 \\ \mathbf{X}'_2 \end{pmatrix}(\mathbf{X}_1, \mathbf{X}_2) = \begin{pmatrix} \mathbf{S}_{11} & \mathbf{S}_{12} \\ \mathbf{S}_{21} & \mathbf{S}_{22} \end{pmatrix},$$

where

$$\mathbf{S}_{11} = \mathbf{X}'_1\mathbf{X}_1, \quad \mathbf{S}_{12} = \mathbf{X}'_1\mathbf{X}_2, \quad \mathbf{S}_{21} = \mathbf{S}'_{12} = \mathbf{X}'_2\mathbf{X}_1, \quad \mathbf{S}_{22} = \mathbf{X}'_2\mathbf{X}_2,$$

$$\mathbf{S}_{11} = \begin{pmatrix} \mathbf{S}_{111} & & \mathbf{0} \\ & \ddots & \\ \mathbf{0} & & \mathbf{S}_{k11} \end{pmatrix}, \quad \mathbf{S}_{12} = \begin{pmatrix} \mathbf{S}_{112} \\ \vdots \\ \mathbf{S}_{k12} \end{pmatrix}, \quad \mathbf{S}_{22} = (\mathbf{S}_{122} + \ldots + \mathbf{S}_{k22}),$$

and where

$$\mathbf{S}_{j11} = \mathbf{X}'_{j1}\mathbf{X}_{j1}, \quad \mathbf{S}_{j12} = \mathbf{X}'_{j1}\mathbf{X}_{j2} = \mathbf{S}'_{j21}, \quad \mathbf{S}_{j22} = \mathbf{X}'_{j2}\mathbf{X}_{j2}.$$

From the normal equations (3·5) we now obtain

$$\left.\begin{aligned} \mathbf{S}_{11}\hat{\boldsymbol{\beta}}_1 + \mathbf{S}_{12}\hat{\boldsymbol{\beta}}_2 &= \mathbf{X}'_1\mathbf{y}, \\ \mathbf{S}_{21}\hat{\boldsymbol{\beta}}_1 + \mathbf{S}_{22}\hat{\boldsymbol{\beta}}_2 &= \mathbf{X}'_2\mathbf{y}, \end{aligned}\right\} \tag{4·2}$$

so that as in § 2 we have

$$\hat{\boldsymbol{\beta}}_2 = \mathbf{S}^{-1}_{22.1}\mathbf{X}'_{2.1}\mathbf{y}, \tag{4·3}$$

where

$$\mathbf{S}_{22.1} = (\mathbf{S}_{22} - \mathbf{S}_{21}\mathbf{S}_{11}^{-1}\mathbf{S}_{12}) = \sum_{j=1}^{k}(\mathbf{S}_{j22} - \mathbf{S}_{j21}\mathbf{S}_{j11}^{-1}\mathbf{S}_{j12}) = \sum_{j=1}^{k}\mathbf{S}_{j22.1},$$

$$\mathbf{X}'_{2.1} = (\mathbf{X}'_{12.1}, \ldots, \mathbf{X}'_{k2.1}) \quad \text{and} \quad \mathbf{X}'_{2.1}\mathbf{y} = \sum_{j=1}^{k}\mathbf{X}'_{j2.1}\mathbf{y}_j.$$

Also as in (2·8), and from the definition of matrices under H_2 we have

$$\begin{aligned} \begin{pmatrix} \hat{\boldsymbol{\beta}}_{11} \\ \vdots \\ \hat{\boldsymbol{\beta}}_{k1} \end{pmatrix} &= \begin{pmatrix} \mathbf{S}_{111}^{-1} & & \mathbf{0} \\ & \ddots & \\ \mathbf{0} & & \mathbf{S}_{k11}^{-1} \end{pmatrix}\left[\begin{pmatrix} \mathbf{X}'_{11} & & \mathbf{0} \\ & \ddots & \\ \mathbf{0} & & \mathbf{X}'_{k1} \end{pmatrix}\begin{pmatrix} \mathbf{y}_1 \\ \vdots \\ \mathbf{y}_k \end{pmatrix} - \begin{pmatrix} \mathbf{S}_{112} \\ \vdots \\ \mathbf{S}_{k12} \end{pmatrix}\hat{\boldsymbol{\beta}}_{.2}\right] \\ &= \begin{pmatrix} \mathbf{S}_{111}^{-1}\mathbf{X}'_{11}\mathbf{y}_1 \\ \vdots \\ \mathbf{S}_{k11}^{-1}\mathbf{X}'_{k1}\mathbf{y}_k \end{pmatrix} - \begin{pmatrix} \mathbf{S}_{111}^{-1}\mathbf{S}_{112}\hat{\boldsymbol{\beta}}_{.2} \\ \vdots \\ \mathbf{S}_{k11}^{-1}\mathbf{S}_{k12}\hat{\boldsymbol{\beta}}_{.2} \end{pmatrix}. \end{aligned} \tag{4·4}$$

Thus under H_2 of (4·1), we have the following estimates of the regression coefficients:

$$\left.\begin{aligned} \hat{\boldsymbol{\beta}}_{j1} &= \mathbf{S}_{j11}^{-1}(\mathbf{X}_{j1}'\mathbf{y}_j - \mathbf{S}_{j12}\hat{\boldsymbol{\beta}}_{.2}) = \hat{\boldsymbol{\beta}}_{j1^2} \quad (j = 1, 2, \ldots, k), \\ \hat{\boldsymbol{\beta}}_{.2} &= \left(\sum_{j=1}^{k} \mathbf{S}_{j22.1}\right)^{-1} \sum_{j=1}^{k} \mathbf{X}_{j2.1}\mathbf{y}_j = \hat{\boldsymbol{\beta}}_{.2^2}, \end{aligned}\right\} \tag{4·5}$$

where $$\hat{\boldsymbol{\beta}}_j' = (\hat{\boldsymbol{\beta}}_{j1}', \hat{\boldsymbol{\beta}}_{.2}') = \hat{\boldsymbol{\beta}}_{j^2}'.$$

If under H_1 of (3·8) we also define $\boldsymbol{\beta}_j' = (\boldsymbol{\beta}_{j1}', \boldsymbol{\beta}_{j2}')$ but rearrange and partition the sub-matrices of $\boldsymbol{\beta}$ and $\mathbf{X}$ so that

$$\boldsymbol{\beta}' = (\boldsymbol{\beta}_1', \boldsymbol{\beta}_2'), \quad \mathbf{X} = (\mathbf{X}_1, \mathbf{X}_2),$$

where $$\boldsymbol{\beta}_1' = (\boldsymbol{\beta}_{11}', \boldsymbol{\beta}_{21}', \ldots, \boldsymbol{\beta}_{k1}'), \quad \boldsymbol{\beta}_{j1}' = (\beta_{j1}, \beta_{j2}, \ldots, \beta_{jq}),$$

$$\boldsymbol{\beta}_2' = (\boldsymbol{\beta}_{12}', \boldsymbol{\beta}_{22}', \ldots, \boldsymbol{\beta}_{k2}'), \quad \boldsymbol{\beta}_{j2}' = (\beta_{jq+1}', \beta_{jq+2}, \ldots, \beta_{jp}),$$

$$\mathbf{X}_1 = \begin{pmatrix} \mathbf{X}_{11} & & \mathbf{0} \\ & \ddots & \\ \mathbf{0} & & \mathbf{X}_{k1} \end{pmatrix}, \quad \mathbf{X}_2 = \begin{pmatrix} \mathbf{X}_{12} & & \mathbf{0} \\ & \ddots & \\ \mathbf{0} & & \mathbf{X}_{k2} \end{pmatrix},$$

$$\mathbf{X}_{j1} = (\mathbf{x}_{j1}, \mathbf{x}_{j2}, \ldots, \mathbf{x}_{jq}), \quad \mathbf{X}_{j2} = (\mathbf{x}_{jq+1}, \mathbf{x}_{jq+2}, \ldots, \mathbf{x}_{jp}) \quad (j = 1, 2, \ldots, k),$$

then $$\mathbf{X}_1'\mathbf{X}_1 = \mathbf{S}_{11} = \begin{pmatrix} \mathbf{S}_{111} & & \mathbf{0} \\ & \ddots & \\ \mathbf{0} & & \mathbf{S}_{k11} \end{pmatrix}, \quad \mathbf{X}_1'\mathbf{X}_2 = \mathbf{S}_{12} = \begin{pmatrix} \mathbf{S}_{112} & & \mathbf{0} \\ & \ddots & \\ \mathbf{0} & & \mathbf{S}_{k12} \end{pmatrix} = \mathbf{S}_{21}',$$

$$\mathbf{X}_2'\mathbf{X}_2 = \mathbf{S}_{22} = \begin{pmatrix} \mathbf{S}_{122} & & \mathbf{0} \\ & \ddots & \\ \mathbf{0} & & \mathbf{S}_{k22} \end{pmatrix}, \quad \mathbf{X}'\mathbf{X} = \mathbf{S} = \begin{pmatrix} \mathbf{S}_{11} & \mathbf{S}_{12} \\ \mathbf{S}_{21} & \mathbf{S}_{22} \end{pmatrix},$$

and $$\mathbf{S}_{j11} = \mathbf{X}_{j1}'\mathbf{X}_{j1}, \quad \mathbf{S}_{j12} = \mathbf{X}_{j1}'\mathbf{X}_{j2} = \mathbf{S}_{j21}' \quad \text{and} \quad \mathbf{S}_{j22} = \mathbf{X}_{j2}'\mathbf{X}_{j2} \quad (j = 1, 2, \ldots, k).$$

We then obtain the same estimate of $\boldsymbol{\beta}_j$ $(j = 1, 2, \ldots, k)$, as in § 3, by the procedure of § 2, that is, from (3·5) we have

$$\left.\begin{aligned} \mathbf{S}_{11}\hat{\boldsymbol{\beta}}_1 + \mathbf{S}_{12}\hat{\boldsymbol{\beta}}_2 &= \mathbf{X}_1'\mathbf{y}, \\ \mathbf{S}_{21}\hat{\boldsymbol{\beta}}_1 + \mathbf{S}_{22}\hat{\boldsymbol{\beta}}_2 &= \mathbf{X}_2'\mathbf{y}, \end{aligned}\right\} \tag{4·6}$$

$$\hat{\boldsymbol{\beta}}_1 = \mathbf{S}_{11}^{-1}(\mathbf{X}_1'\mathbf{y} - \mathbf{S}_{12}\hat{\boldsymbol{\beta}}_2), \tag{4·7}$$

$$\hat{\boldsymbol{\beta}}_2 = \mathbf{S}_{22.1}^{-1}\mathbf{X}_{2.1}'\mathbf{y}, \tag{4·8}$$

where $$\mathbf{S}_{22.1} = (\mathbf{S}_{21} - \mathbf{S}_{21}\mathbf{S}_{11}^{-1}\mathbf{S}_{12}) = \begin{pmatrix} \mathbf{S}_{122.1} & & \mathbf{0} \\ & \ddots & \\ \mathbf{0} & & \mathbf{S}_{k22.1} \end{pmatrix},$$

$$\mathbf{S}_{j22.1} = (\mathbf{S}_{j22} - \mathbf{S}_{j21}\mathbf{S}_{j11}^{-1}\mathbf{S}_{j12}) \quad (j = 1, 2, \ldots, k),$$

$$\mathbf{X}_{2.1}' = \mathbf{X}_2' - \mathbf{S}_{21}\mathbf{S}_{11}^{-1}\mathbf{X}_1' = \begin{pmatrix} \mathbf{X}_{12.1}' & & \mathbf{0} \\ & \ddots & \\ \mathbf{0} & & \mathbf{X}_{k2.1}' \end{pmatrix},$$

$$\mathbf{X}_{j2.1}' = \mathbf{X}_{j2}' - \mathbf{S}_{j21}\mathbf{S}_{j11}^{-1}\mathbf{X}_{j1}' \quad (j = 1, 2, \ldots, k),$$

so that from (4·7) we obtain under H_1 for each category $j = 1, 2, \ldots, k$,

$$\hat{\boldsymbol{\beta}}_{j1} = \mathbf{S}_{j11}^{-1}(\mathbf{X}'_{j1}\mathbf{y}_j - \mathbf{S}_{j12}\hat{\boldsymbol{\beta}}_{j2}) = \hat{\boldsymbol{\beta}}_{j1}{}^1, \tag{4·9}$$

$$\hat{\boldsymbol{\beta}}_{j2} = \mathbf{S}_{j22.1}^{-1}\mathbf{X}'_{j2.1}\mathbf{y}_j = \hat{\boldsymbol{\beta}}_{j2}{}^1, \tag{4·10}$$

and
$$\hat{\boldsymbol{\beta}}'_j = (\hat{\boldsymbol{\beta}}'_{j1}, \hat{\boldsymbol{\beta}}'_{j2}) = \hat{\boldsymbol{\beta}}'_j{}^1.$$

With these estimates of the parameters under H_1 of (3·8) and H_2 of (4·1) and noting after some reduction that

$$\hat{\boldsymbol{\beta}}'^{\alpha}\mathbf{S}^{\alpha}\boldsymbol{\beta}^{\alpha} = \mathbf{y}'\mathbf{X}_1\mathbf{S}_{11}^{-1}\mathbf{X}'_1\mathbf{y} + \hat{\boldsymbol{\beta}}'^{\alpha}_2\mathbf{S}_{22.1}{}^{\alpha}\hat{\boldsymbol{\beta}}_2{}^{\alpha} \quad (\alpha = 1, 2), \tag{4·11}$$

we obtain

$$\hat{\sigma}^2\hat{J}(1, 2) = \hat{\boldsymbol{\beta}}'^1_2\mathbf{S}_{22.11}\hat{\boldsymbol{\beta}}_2{}^1 - \hat{\boldsymbol{\beta}}'^2_2\mathbf{S}_{22.1}{}^2\hat{\boldsymbol{\beta}}_2{}^2 = \sum_{j=1}^{k}\hat{\boldsymbol{\beta}}'_{j2}{}^1\mathbf{S}_{j22.1}\hat{\boldsymbol{\beta}}_{j2}{}^1 - \hat{\boldsymbol{\beta}}'^2_2\mathbf{S}_{22.1}{}^2\hat{\boldsymbol{\beta}}_2{}^2, \tag{4·12}$$

where for computational convenience we may write

$$\hat{\boldsymbol{\beta}}'^1_2\mathbf{S}_{22.1}{}^1\hat{\boldsymbol{\beta}}_2{}^1 = \hat{\boldsymbol{\beta}}'^1_2\mathbf{X}'_{2.1}{}^1\mathbf{y} = \sum_{j=1}^{k}\hat{\boldsymbol{\beta}}'_{j2}{}^1\mathbf{X}'_{j2.1}\mathbf{y}, \tag{4·13}$$

$$\hat{\boldsymbol{\beta}}'^2_2\mathbf{S}_{22.1}{}^2\hat{\boldsymbol{\beta}}_2{}^2 = \hat{\boldsymbol{\beta}}'^2_2\mathbf{X}'_{2.1}{}^2\mathbf{y}. \tag{4·14}$$

We may therefore summarize in the analysis of variance Table 4·1, where

$$\hat{J}(1, 2) = (p-q)(k-1)F,$$

and F has the analysis of variance distribution with $(p-q)(k-1)$ and $N-pk$ degrees of freedom when the null hypothesis H_2 of (4·1) is true.

Table 4·1

Variation due to	D.F.	Sum of squares
H_2: $\boldsymbol{\beta}_{j1}{}^2$, $\boldsymbol{\beta}_{.2}{}^2$	$qk+p-q$	$\hat{\boldsymbol{\beta}}'^2\mathbf{S}^2\hat{\boldsymbol{\beta}}^2$
Diff.	$(p-q)(k-1)$	$\hat{\boldsymbol{\beta}}'^1\mathbf{S}^1\hat{\boldsymbol{\beta}}^1 - \hat{\boldsymbol{\beta}}'^2\mathbf{S}^2\hat{\boldsymbol{\beta}}^2 = \sum_{j=1}^{k}\hat{\boldsymbol{\beta}}'_{j2}{}^1\mathbf{S}_{j22.1}\hat{\boldsymbol{\beta}}_{j2}{}^1 - \hat{\boldsymbol{\beta}}'^2_2\mathbf{S}_{22.1}{}^2\hat{\boldsymbol{\beta}}_2{}^2 = \hat{\sigma}^2\hat{J}(1, 2)$
H_1: $\boldsymbol{\beta}_{j1}{}^1$, $\boldsymbol{\beta}_{j2}{}^1$	pk	$\hat{\boldsymbol{\beta}}'^1\mathbf{S}^1\hat{\boldsymbol{\beta}}^1$
Diff.	$N-pk$	$(N-pk)\,\hat{\sigma}^2 = \mathbf{y}'\mathbf{y} - \hat{\boldsymbol{\beta}}'^1\mathbf{S}^1\hat{\boldsymbol{\beta}}^1$
Total	$N = n_1+n_2+\ldots+n_k$	$\mathbf{y}'\mathbf{y}$

4·2. *Three-partition sub-hypothesis.* If the sub-hypothesis requires partitioning the matrices $\mathbf{X}$ and $\boldsymbol{\beta}$ into three submatrices, so that

$$\boldsymbol{\beta}' = (\boldsymbol{\beta}'_1, \boldsymbol{\beta}'_2, \boldsymbol{\beta}'_3) \quad \text{and} \quad \mathbf{X} = (\mathbf{X}_1, \mathbf{X}_2, \mathbf{X}_3),$$

we obtain from $\mathbf{S}\hat{\boldsymbol{\beta}} = \mathbf{X}'\mathbf{y}$ the solutions

$$\left.\begin{aligned}\hat{\boldsymbol{\beta}}_3 &= \mathbf{S}_{33.12}^{-1}\mathbf{X}'_{3.12}\mathbf{y},\\ \hat{\boldsymbol{\beta}}_2 &= \mathbf{S}_{22.1}^{-1}(\mathbf{X}'_{2.1}\mathbf{y} - \mathbf{S}_{23.1}\hat{\boldsymbol{\beta}}_3),\\ \hat{\boldsymbol{\beta}}_1 &= \mathbf{S}_{11}^{-1}(\mathbf{X}'_1\mathbf{y} - \mathbf{S}_{12}\hat{\boldsymbol{\beta}}_2 - \mathbf{S}_{13}\hat{\boldsymbol{\beta}}_3),\end{aligned}\right\} \tag{4·15}$$

where $$\mathbf{S} = \begin{pmatrix} \mathbf{S}_{11} & \mathbf{S}_{12} & \mathbf{S}_{13} \\ \mathbf{S}_{21} & \mathbf{S}_{22} & \mathbf{S}_{23} \\ \mathbf{S}_{31} & \mathbf{S}_{32} & \mathbf{S}_{33} \end{pmatrix}, \quad \mathbf{S}_{tu} = \mathbf{X}_t'\mathbf{X}_u \quad (t, u = 1, 2, 3),$$

and $$\mathbf{S}_{33.12} = \mathbf{S}_{33.1} - \mathbf{S}_{32.1}\mathbf{S}_{22.1}^{-1}\mathbf{S}_{23.1},$$

$$\mathbf{S}_{33.1} = \mathbf{S}_{33} - \mathbf{S}_{31}\mathbf{S}_{11}^{-1}\mathbf{S}_{13}, \qquad \mathbf{S}_{32.1} = \mathbf{S}_{32} - \mathbf{S}_{31}\mathbf{S}_{11}^{-1}\mathbf{S}_{12} = \mathbf{S}_{23.1}',$$

$$\mathbf{S}_{22.1} = \mathbf{S}_{22} - \mathbf{S}_{21}\mathbf{S}_{11}^{-1}\mathbf{S}_{12}, \qquad \mathbf{X}_{2.1}'\mathbf{y} = (\mathbf{X}_2' - \mathbf{S}_{21}\mathbf{S}_{11}^{-1}\mathbf{X}_1')\mathbf{y},$$

$$\mathbf{X}_{3.12}'\mathbf{y} = (\mathbf{X}_{3.1}' - \mathbf{S}_{32.1}\mathbf{S}_{22.1}^{-1}\mathbf{X}_{2.1}')\mathbf{y}, \quad \mathbf{X}_{3.1}'\mathbf{y} = (\mathbf{X}_3' - \mathbf{S}_{31}\mathbf{S}_{11}^{-1}\mathbf{X}_1')\mathbf{y}.$$

Also (compare with (4·11))

$$\hat{\boldsymbol{\beta}}'\mathbf{S}\hat{\boldsymbol{\beta}} = \mathbf{y}'\mathbf{X}_1\mathbf{S}_{11}^{-1}\mathbf{X}_1'\mathbf{y} + \mathbf{y}'\mathbf{X}_{2.1}\mathbf{S}_{22.1}^{-1}\mathbf{X}_{2.1}'\mathbf{y} + \hat{\boldsymbol{\beta}}_3'\mathbf{S}_{33.12}\hat{\boldsymbol{\beta}}_3. \tag{4·16}$$

Using (4·15) and collecting terms we obtain other useful forms of equation (4·16), e.g.

$$\hat{\boldsymbol{\beta}}'\mathbf{S}\hat{\boldsymbol{\beta}} = \mathbf{y}'\mathbf{X}_1\mathbf{S}_{11}^{-1}\mathbf{X}_1'\mathbf{y} + \hat{\boldsymbol{\beta}}_2'\mathbf{X}_{2.1}'\mathbf{y} + \hat{\boldsymbol{\beta}}_3'\mathbf{X}_{3.1}'\mathbf{y}, \tag{4·16·1}$$

which is convenient when the data are raw observations and $x_{ji1} = 1$ for all j and i, so that the first partition serves to include the x_{ji1} and β_{j1} and obtain deviations about average values for basically a two-partition problem, and

$$\hat{\boldsymbol{\beta}}'\mathbf{S}\hat{\boldsymbol{\beta}} = \hat{\boldsymbol{\beta}}_1'\mathbf{X}_1'\mathbf{y} + \hat{\boldsymbol{\beta}}_2'\mathbf{X}_2'\mathbf{y} + \hat{\boldsymbol{\beta}}_3'\mathbf{X}_3'\mathbf{y} \tag{4·16·2}$$

for a three-partition problem where the variables already are deviations about average values.

The above discussion, if required, can readily be extended by induction to any number of partitions.

4·3. *Carter's regression case.* Carter (1949) considers the case of a correlation effect among the ith observations ($i = 1, 2, \ldots, n$) in each of k samples. His regression model can be written as

$$z_{ji} = y_{ji} - \sum_{r=1}^{q} \beta_{jr}x_{jir} - \alpha_i, \tag{4·17}$$

where the correlation effect among samples is due to α_i, an element common to the ith observation in each sample, $j = 1, 2, \ldots, k$.

It can be seen that this model is a particular case of the sub-hypothesis analysis where the $\mathbf{X}$ and $\boldsymbol{\beta}$ matrices are $$\boldsymbol{\beta}' = (\boldsymbol{\beta}_1', \boldsymbol{\beta}_2'), \quad \mathbf{X} = (\mathbf{X}_1, \mathbf{X}_2),$$

and the submatrices are

$$\boldsymbol{\beta}_1' = (\boldsymbol{\beta}_{11}'\boldsymbol{\beta}_{21}', \ldots, \boldsymbol{\beta}_{k1}'), \quad \boldsymbol{\beta}_{j1}' = (\beta_{j1}, \beta_{j2}, \ldots, \beta_{jq}), \quad \boldsymbol{\beta}_2' = (\alpha_1, \alpha_2, \ldots, \alpha_n),$$

$$\mathbf{X}_1 = \begin{pmatrix} \mathbf{X}_{11} & & \mathbf{0} \\ & \ddots & \\ \mathbf{0} & & \mathbf{X}_{k1} \end{pmatrix}, \quad \mathbf{X}_{j1} = \begin{pmatrix} x_{j11} & \cdots & x_{j1q} \\ \vdots & & \vdots \\ x_{jn1} & \cdots & x_{jnq} \end{pmatrix},$$

$$\mathbf{X}_2 = \begin{pmatrix} \mathbf{I} \\ \vdots \\ \mathbf{I} \end{pmatrix}, \qquad \mathbf{I} = \begin{pmatrix} 1 & & 0 \\ & \ddots & \\ 0 & & 1 \end{pmatrix},$$

where $\mathbf{X}_2$ is a $k \times 1$ matrix of submatrices $\mathbf{I}$, the identity matrix of order $n \times n$. With these definitions of $\mathbf{X}$ and $\boldsymbol{\beta}$, Carter's normal equations (his (3·3)) for estimating the β's follow directly from the normal equations (4·2) by obtaining

$$\mathbf{S}_{11.2}\hat{\boldsymbol{\beta}}_1 = \mathbf{X}'_{1.2}\mathbf{y}, \tag{4·18}$$

where $$\mathbf{S}_{11.2} = \mathbf{S}_{11} - \mathbf{S}_{12}\mathbf{S}_{22}^{-1}\mathbf{S}_{21}.$$

For this case we obtain

$$\mathbf{S}_{11.2} = \begin{pmatrix} \left(1-\frac{1}{k}\right)\mathbf{X}'_{11}\mathbf{X}_{11} & -\frac{1}{k}\mathbf{X}'_{11}\mathbf{X}_{21} & \dots & -\frac{1}{k}\mathbf{X}'_{11}\mathbf{X}_{k1} \\ -\frac{1}{k}\mathbf{X}'_{21}\mathbf{X}_{11} & \left(1-\frac{1}{k}\right)\mathbf{X}'_{21}\mathbf{X}_{21} & \dots & -\frac{1}{k}\mathbf{X}'_{21}\mathbf{X}_{k1} \\ \vdots & \vdots & & \vdots \\ -\frac{1}{k}\mathbf{X}'_{k1}\mathbf{X}_{11} & -\frac{1}{k}\mathbf{X}'_{k1}\mathbf{X}_{21} & \dots & \left(1-\frac{1}{k}\right)\mathbf{X}'_{k1}\mathbf{X}_{k1} \end{pmatrix},$$

and

$$\mathbf{X}'_{1.2} = \begin{pmatrix} \left(1-\frac{1}{k}\right)\mathbf{X}'_{11} & -\frac{1}{k}\mathbf{X}'_{11} & \dots & -\frac{1}{k}\mathbf{X}'_{11} \\ -\frac{1}{k}\mathbf{X}'_{21} & \left(1-\frac{1}{k}\right)\mathbf{X}'_{21} & \dots & -\frac{1}{k}\mathbf{X}'_{21} \\ \vdots & \vdots & & \vdots \\ -\frac{1}{k}\mathbf{X}'_{k1} & -\frac{1}{k}\mathbf{X}'_{k1} & \dots & \left(1-\frac{1}{k}\right)\mathbf{X}'_{k1} \end{pmatrix}.$$

As before, $$\mathbf{S}_{11} = \mathbf{X}'_1\mathbf{X}_1, \quad \mathbf{S}_{12} = \mathbf{X}'_1\mathbf{X}_2 = \mathbf{S}'_{21}, \quad \mathbf{S}_{22} = \mathbf{X}'_2\mathbf{X}_2.$$

The estimates of the correlation effects α_i are not given specifically by Carter. The solution

$$\hat{\alpha}_i = \bar{z}_i \quad (i = 1, 2, \dots, n), \tag{4·19}$$

where $$\bar{z}_i = \frac{1}{k}\sum_{j=1}^{k}\hat{z}_{ji} = \frac{1}{k}\sum_{j=1}^{k}\left(y_{ji} - \sum_{r=1}^{q}\hat{\beta}_{jr}x_{jir}\right)$$

follows directly from $$\mathbf{S}_{22}\hat{\boldsymbol{\beta}}_2 = \mathbf{X}'_2\mathbf{y} - \mathbf{S}_{21}\hat{\boldsymbol{\beta}}_1. \tag{4·20}$$

5. Example

As an example of the foregoing results, consider the performance data of a manufactured product tested under three environmental conditions (categories) each involving three independent variables. In the equation

$$z_{ji} = y_{ji} - \beta_{j1}x_{ji1} - \beta_{j2}x_{ji2} - \beta_{j3}x_{ji3} - \beta_{j4}x_{ji4}, \tag{5·1}$$

the data y_{ji} and x_{jir} $(r = 1, 2, 3, 4)$ are raw observations, so that $x_{ji1} = 1$ for all $j = 1, 2, 3$ and $i = 1, 2, \dots, n_j$. In this example $k = 3, p = 4, n_1 = 16, n_2 = 15$ and $n_3 = 16$. The matrices

$\mathbf{S}_j$ and $\mathbf{X}_j'\mathbf{y}_j$ $(j = 1, 2, 3)$ of the computed sums of squares and products about the origin are

$$\mathbf{S}_1 = \begin{pmatrix} 16{\cdot}0 & 286{\cdot}8 & 139{\cdot}0 & 4{,}835{\cdot}0 \\ 286{\cdot}8 & 5{,}340{\cdot}4 & 2{,}452{\cdot}2 & 86{,}849{\cdot}0 \\ 139{\cdot}0 & 2{,}452{\cdot}2 & 1{,}307{\cdot}0 & 41{,}990{\cdot}0 \\ 4{,}835{\cdot}0 & 86{,}849{\cdot}0 & 41{,}990{\cdot}0 & 1{,}465{,}575{\cdot}0 \end{pmatrix}, \quad \mathbf{X}_1'\mathbf{y}_1 = \begin{pmatrix} 97{,}500 \\ 1{,}788{,}052 \\ 838{,}010 \\ 29{,}484{,}809 \end{pmatrix},$$

$$\mathbf{S}_2 = \begin{pmatrix} 15{\cdot}0 & 244{\cdot}6 & 236{\cdot}0 & 4{,}625{\cdot}0 \\ 244{\cdot}6 & 4{,}181{\cdot}6 & 3{,}869{\cdot}0 & 75{,}318{\cdot}0 \\ 236{\cdot}0 & 3{,}869{\cdot}0 & 3{,}824{\cdot}0 & 72{,}500{\cdot}0 \\ 4{,}625{\cdot}0 & 75{,}318{\cdot}0 & 72{,}500{\cdot}0 & 1{,}427{,}425{\cdot}0 \end{pmatrix}, \quad \mathbf{X}_2'\mathbf{y}_2 = \begin{pmatrix} 83{,}470 \\ 1{,}404{,}814 \\ 1{,}320{,}100 \\ 25{,}727{,}050 \end{pmatrix},$$

$$\mathbf{S}_3 = \begin{pmatrix} 16{\cdot}0 & 256{\cdot}0 & 97{\cdot}0 & 2{,}995{\cdot}0 \\ 256{\cdot}0 & 4{,}221{\cdot}7 & 1{,}619{\cdot}2 & 47{,}897{\cdot}0 \\ 97{\cdot}0 & 1{,}619{\cdot}2 & 785{\cdot}0 & 17{,}840{\cdot}0 \\ 2{,}995{\cdot}0 & 47{,}897{\cdot}0 & 17{,}840{\cdot}0 & 580{,}475{\cdot}0 \end{pmatrix}, \quad \mathbf{X}_3'\mathbf{y}_3 = \begin{pmatrix} 89{,}280 \\ 1{,}456{,}596 \\ 554{,}650 \\ 16{,}743{,}450 \end{pmatrix},$$

where

$$\mathbf{S}_j = (s_{jrt}), \quad s_{jrt} = \sum_{i=1}^{n_j} x_{jir}x_{jit} \quad (r, t = 1, 2, 3, 4) \quad \text{and} \quad \mathbf{X}_j'\mathbf{y}_j = (s_{jyx_r}), \quad s_{jyx_r} = \sum_{i=1}^{n_j} y_{ji}x_{jir}.$$

It may be noted in the above that the element $s_{111} = n_1$, $s_{211} = n_2$, and $s_{311} = n_3$. The multiple regression equation for all three categories combined is given by (3·4), where it will be remembered that specification of the matrices $\mathbf{X}^\alpha$ and $\boldsymbol{\beta}^\alpha$ depends on the model prescribed by hypothesis. The data in the above matrices can be suitably arranged for analysis according to hypothesis.

To illustrate the statistical method seven hypotheses are considered and tested. The hypothesis H_1 imposes no restriction on the β's so that

$$H_1: \quad \beta_{j1} = \beta_{j1}, \quad \beta_{jr} = \beta_{jr} \quad (r = 2, 3, 4).$$

All other hypotheses are compared as null hypotheses against H_1:

$$\begin{aligned}
H_2&: \quad \beta_{j1} = \beta_{j1}, \quad \beta_{jr} = 0 \quad (r = 2, 3, 4), \\
H_3&: \quad \beta_{j1} = \beta_{.1}, \quad \beta_{jr} = \beta_{.r} \quad (r = 2, 3, 4), \\
H_4&: \quad \beta_{j1} = \beta_{j1}, \quad \beta_{jr} = \beta_{.r} \quad (r = 2, 3, 4), \\
H_5&: \quad \beta_{j1} = \beta_{j1}, \quad \beta_{jr} = \beta_{jr} \quad (r = 2); \qquad \beta_{jr} = \beta_{.r} \quad (r = 3, 4), \\
H_6&: \quad \beta_{j1} = \beta_{j1}, \quad \beta_{jr} = \beta_{jr} \quad (r = 2); \qquad \beta_{jr} = 0 \quad (r = 3, 4), \\
H_7&: \quad \beta_{j1} = \beta_{j1}, \quad \beta_{jr} = \beta_{.r} \quad (r = 2); \qquad \beta_{jr} = \beta_{jr} \quad (r = 3, 4).
\end{aligned}$$

The above statements of the various hypotheses all apply for $j = 1, 2, 3$. In stating these hypotheses we have specified β_{j1} separately, for convenience, since in this example it represents the constant term which depends on the mean values. Table 5·1 presents the complete summary of the analysis of variance data and the tests of significance of the various hypotheses. Table 5·2 presents the estimated regression coefficients of the various hypotheses. The specification of the matrices $\mathbf{X}^\alpha$ and $\boldsymbol{\beta}^\alpha$ for H_1 and H_5 is also given, following Table 5·2; those for the other hypotheses follow on the same lines.

Table 5·1. *Analysis of variance table for tests of various null hypotheses* H_α ($\alpha = 2, 3, 4, 5, 6, 7$) *against an alternative hypothesis* H_1

Variation due to	D.F.	Sum of squares		$\hat{J}(1, \alpha)$	F
H_2: $\beta_{j1} = \beta_{j1}$ $\beta_{jr} = 0$ $(r = 2, 3, 4)$ $\boldsymbol{\beta}^2$	$k =$ 3		1,556,805,752		
Diff. H_1, H_2	$pk-k =$ 9		24,992,970	870·41	96·7**
H_3: $\beta_{j1} = \beta_{.1}$	1 1		1,553,937,500		
$\beta_{jr} = \beta_{.r}$ $(r = 2, 3, 4)$	$p-1$ 3		27,027,426		
$\boldsymbol{\beta}^3$	p 4		1,580,964,926		
Diff. H_1, H_3	$pk-p$ 8		833,796	29·04	3·63**
H_4: $\beta_{j1} = \beta_{j1}$	k 3		1,556,805,752		
$\beta_{jr} = \beta_{.r}$ $(r = 2, 3, 4)$	$p-1$ 3		24,325,402		
$\boldsymbol{\beta}^4$	$p+k-1$ 6		1,581,131,154		
Diff. H_1, H_4	$(p-1)(k-1)$ 6		667,568	23·25	3·88**
H_5: $\beta_{j1} = \beta_{j1}$	k 3	$\mathbf{y}'\mathbf{X}_1\mathbf{S}_{11}^{-1}\mathbf{X}_1'\mathbf{y}$	1,556,805,752		
$\beta_{jr} = \beta_{jr}$ $(r = 2)$	k 3	$\hat{\boldsymbol{\beta}}_2'\mathbf{X}_{2.1}'\mathbf{y}$	23,923,966		
$= \beta_{.r}$ $(r = 3, 4)$	$p-2$ 2	$\hat{\boldsymbol{\beta}}_3'\mathbf{X}_{3.1}'\mathbf{y}$	493,550		
$\boldsymbol{\beta}^5$	$p+2(k-1)$ 8	$\hat{\boldsymbol{\beta}}'^5\mathbf{S}^5\hat{\boldsymbol{\beta}}^5$	1,581,223,268		
Diff. H_1, H_5	$(p-2)(k-1)$ 4	$\hat{\boldsymbol{\beta}}'^1\mathbf{S}^1\hat{\boldsymbol{\beta}}^1 - \hat{\boldsymbol{\beta}}'^5\mathbf{S}^5\hat{\boldsymbol{\beta}}^5 = \hat{\sigma}^2\hat{J}(1,5)$	575,454	20·04	5·01**
H_6: $\beta_{j1} = \beta_{j1}$	k 3		1,556,805,752		
$\beta_{jr} = \beta_{jr}$ $(r = 2)$	k 3		24,349,074		
$= 0$ $(r = 3, 4)$					
$\boldsymbol{\beta}^6$	$2k$ 6		1,581,154,826		
Diff. H_1, H_6	$(p-2)k$ 6		643,896	22·42	3·74**
H_7: $\beta_{j1} = \beta_{j1}$	k 3		1,556,805,752		
$\beta_{jr} = \beta_{.r}$ $(r = 2)$	1 1		2,450,676		
$= \beta_{jr}$ $(r = 3, 4)$	$2(p-1)$ 6		22,461,479		
$\boldsymbol{\beta}^7$	$2p+k-1$ 10		1,581,717,907		
Diff. H_1, H_7	$p(k-2)-(k-1)$ 2		80,815	2·81	1·41
H_1: $\beta_{j1} = \beta_{j1}$	k 3	$\mathbf{y}'\mathbf{X}_1\mathbf{S}_{11}^{-1}\mathbf{X}_1'\mathbf{y}$	1,556,805,752		
$\beta_{jr} = \beta_{jr}$ $(r = 2, 3, 4)$	$pk-k$ 9	$\hat{\boldsymbol{\beta}}_2'\mathbf{X}_{2.1}'\mathbf{y}$	24,992,970		
$\boldsymbol{\beta}^1$	pk 12	$\hat{\boldsymbol{\beta}}'^1\mathbf{S}^1\hat{\boldsymbol{\beta}}^1$	1,581,798,722		
Diff.	$N-pk$ 35	$\mathbf{y}'\mathbf{y} - \hat{\boldsymbol{\beta}}'^1\mathbf{S}^1\hat{\boldsymbol{\beta}}^1 = (N-pk)\,\hat{\sigma}^2$	1,004,984	$\hat{\sigma}^2 = 28{,}714$	
Total	$N = n_1+n_2+n_3$ 47	$\mathbf{y}'\mathbf{y}$	1,582,803,706		

** Significance at 0·01 probability level.

Using the 0·01 probability level for significance, and the 0·05 probability level for caution, it is concluded, from Table 5·1, that

(1) The regression is real; reject H_2.

(2) One set of regression coefficients, including equality of means, cannot adequately represent all three categories; reject H_3.

(3) One set of regression coefficients is not adequate even after allowing for differences in mean value for each category; reject H_4.

(4) One set of regression coefficients for variables x_3 and x_4 for all three categories cannot be used; reject H_5, further

(5) the regression coefficients for x_3 and x_4 cannot be ignored; reject H_6.

However,

(6) the use of one regression coefficient for the variable x_2 and different ones for x_3 and x_4 and for the constant term is adequate; accept H_7.

Table 5·2. *Estimates of regression coefficients under various hypotheses*

Hypothesis	β_{j1}	β_{j2}	β_{j3}	β_{j4}
H_1: $j=1$	3586	203·4	−10·69	−3·46
$j=2$	−7186	231·1	79·02	25·10
$j=3$	1654	227·7	−6·93	1·73
H_2: $j=1$	6094			
$j=2$	5564			
$j=3$	5580			
H_3: $j=1, 2, 3$	2009	219·8	−11·19	0·647
H_4: $j=1$	1803	216·0	3·71	1·28
$j=2$	1589	216·0	3·71	1·28
$j=3$	1862	216·0	3·71	1·28
H_5: $j=1$	1349	206·9	34·64	2·43
$j=2$	617	224·0	34·64	2·43
$j=3$	1625	205·6	34·64	2·43
H_6: $j=1$	2467	202·3		
$j=2$	1873	226·4		
$j=3$	2001	223·7		
H_7: $j=1$	3431	219·8	−4·28	−4·10
$j=2$	−6767	219·8	79·26	24·33
$j=3$	1758	219·8	−4·15	1·77

For the hypotheses H_1 and H_5 considered in the example the matrix of parameters $\boldsymbol{\beta}$, and the matrix of observations $\mathbf{X}$ are given below. It will be noted that since we are dealing with raw observations in the example, the regression coefficients β_{j1} of $\boldsymbol{\beta}$ and the vector $\mathbf{x}_{j1}$ of $\mathbf{X}$ ($j = 1, 2, 3$) were partitioned for every hypothesis. This provides for the usual practice of obtaining sums of squares and products of deviations about average values to

simplify further calculations by reducing by one the rank of matrices **S** (of sums of squares and products) whose inverses must be obtained.

$$H_1: \quad \beta_{j1} = \beta_{j1}, \quad \beta_{jr} = \beta_{jr} \quad (r = 2, 3, 4;\ j = 1, 2, 3),$$

$$\boldsymbol{\beta}' = (\boldsymbol{\beta}_1', \boldsymbol{\beta}_2'), \quad \mathbf{X} = (\mathbf{X}_1, \mathbf{X}_2),$$

$$\boldsymbol{\beta}_1' = (\beta_{11}, \beta_{21}, \beta_{31}), \quad \boldsymbol{\beta}_2' = (\boldsymbol{\beta}_{12}', \boldsymbol{\beta}_{22}', \boldsymbol{\beta}_{32}'), \quad \boldsymbol{\beta}_{j2}' = (\beta_{j2}, \beta_{j3}, \beta_{j4}),$$

$$\mathbf{X}_1 = \begin{pmatrix} \mathbf{x}_{11} & . & \mathbf{0} \\ . & \mathbf{x}_{21} & . \\ \mathbf{0} & . & \mathbf{x}_{31} \end{pmatrix}, \quad \mathbf{X}_2 = \begin{pmatrix} \mathbf{X}_{12} & . & \mathbf{0} \\ . & \mathbf{X}_{22} & . \\ \mathbf{0} & . & \mathbf{X}_{32} \end{pmatrix},$$

$$\mathbf{x}_{j1}' = (1, 1, \ldots, 1) \qquad \mathbf{X}_{j2} = (\mathbf{x}_{j2}, \mathbf{x}_{j3}, \mathbf{x}_{j4}),$$

$$\text{order } 1 \times n_j, \qquad \mathbf{x}_{jr}' = (x_{j1r}, x_{j2r}, \ldots, x_{jn_jr}),$$

$$H_5: \quad \beta_{j1} = \beta_{j1}, \quad \beta_{jr} = \beta_{jr} \quad (r = 2), \qquad \beta_{jr} = \beta_{.r} \quad (r = 3, 4;\ j = 1, 2, 3),$$

$$\boldsymbol{\beta}' = (\boldsymbol{\beta}_1', \boldsymbol{\beta}_2', \boldsymbol{\beta}_3'), \quad \mathbf{X} = (\mathbf{X}_1, \mathbf{X}_2, \mathbf{X}_3),$$

$$\boldsymbol{\beta}_1' = (\beta_{11}, \beta_{21}, \beta_{31}), \quad \boldsymbol{\beta}_2' = (\beta_{12}, \beta_{22}, \beta_{32}), \quad \boldsymbol{\beta}_3' = (\beta_{.3}, \beta_{.4}),$$

$$\mathbf{X}_1 = \begin{pmatrix} \mathbf{x}_{11} & . & \mathbf{0} \\ . & \mathbf{x}_{21} & . \\ \mathbf{0} & . & \mathbf{x}_{31} \end{pmatrix}, \quad \mathbf{X}_2 = \begin{pmatrix} \mathbf{x}_{12} & . & \mathbf{0} \\ . & \mathbf{x}_{22} & . \\ \mathbf{0} & . & \mathbf{x}_{32} \end{pmatrix}, \quad \begin{matrix} \mathbf{X}_3' = (\mathbf{X}_{13}', \mathbf{X}_{23}', \mathbf{X}_{33}'), \\ \mathbf{X}_{j3} = (\mathbf{x}_{j3}, \mathbf{x}_{j4}), \end{matrix}$$

$\mathbf{x}_{j1}$ and $\mathbf{x}_{jr}$ $(r = 2, 3, 4)$, are defined as under H_1.

In the foregoing example each hypothesis on the parameters applied to all categories, $j = 1, 2, 3$. It is clear, however, that this need not be the case for the theory and method are equally applicable for any assertion of the hypotheses with regard to the parameters. For example, we might have considered a case where part of the hypothesis concerned equality of the parameters for certain of the categories, but not for all, e.g.

$$\begin{aligned} H_8: \quad & \beta_{jr} = \beta_{.r} \quad (j = 1, 3; \quad r = 1), \\ & \beta_{jr} = \beta_{jr} \quad (j = 2; \quad r = 1), \\ & \beta_{jr} = \beta_{.r} \quad (j = 1, 2, 3;\ r = 2), \\ & \beta_{jr} = \beta_{jr} \quad (j = 1, 2, 3;\ r = 3, 4), \end{aligned}$$

and analysis by the three-partition sub-hypothesis procedure of § 4 would apply.

The authors wish to express their sincere appreciation to the computing staff at the Naval Proving Ground, Dahlgren, Virginia, and to Mr Fred Okano for their able assistance in carrying out the computations.

REFERENCES

Carter, A. H. (1949). The estimation and comparison of residual regressions where there are two or more related sets of observations. *Biometrika*, **36**, 26–46.

Frazer, R. A., Duncan, W. J. & Collar, A. R. (1938). *Elementary Matrices*. Cambridge University Press.

Grundy, P. M. (1951). A general technique for the analysis of experiments with incorrectly treated plots. *J. R. Statist. Soc.* B, **13**, 272–83.

Kempthorne, O. (1952). *The Design and Analysis of Experiments*. New York: John Wiley and Sons, Inc.

KENDALL, M. G. (1946). *The Advanced Theory of Statistics*, **2**. London: Chas. Griffin and Co., Ltd.
KOLODZIEJCZYK, S. (1935). On an important class of statistical hypotheses. *Biometrika*, **27**, 161–90.
KULLBACK, S. (1952). An application of information theory to multivariate analysis. *Ann. Math. Statist.* **23**, 88–102.
KULLBACK, S. (1956). An application of information theory to multivariate analysis. II. *Ann. Math. Statist.* **27**, 122–46.
KULLBACK, S. & LEIBLER, R. A. (1951). On information and sufficiency. *Ann. Math. Statist.* **22**, 79–86.
PEARSON, E. S. & WILKS, S. S. (1933). Methods of statistical analysis appropriate for k samples of two variables. *Biometrika*, **25**, 353–78.
PLACKETT, R. L. (1949). A historical note on the method of least squares. *Biometrika*, **36**, 458–60.
RAO, C. R. (1952). *Advanced Statistical Methods in Biometric Research*. New York: John Wiley and Sons, Inc.
ROSENBLATT, H. M. (1953). On a k sample multivariate regression problem. Master's Thesis, The George Washington University.
WELCH, B. L. (1935). Problems in the analysis of regression among k samples. *Biometrika*, **27**, 145–60.
WILKS, S. S. (1943). *Mathematical Statistics*. Princeton University Press.
WILLIAMS, E. J. (1953). Tests of significance for concurrent regression lines. *Biometrika*, **40**, 297–305.

BIVARIATE STRUCTURAL RELATION

By R. L. BROWN

British Coal Utilization Research Association, Leatherhead, Surrey

SUMMARY. Given n observations (x_i, y_i), each subject to an error of measurement *ascertained independently*, the existence of a structural relation for the true points (X, Y) is discussed. For a linear relation a ϕ-number is defined from which either a confidence or a *structural* theory can be developed. The relation is treated *in toto*. Structural theory is described in terms of three hypotheses of regularity, which enable confidence and fiducial theories to be contrasted. It is shown that the theory provides a solution for the curvilinear relation wherein all the alternative hypotheses are incorporated, and it is suggested that the latter is a necessary requirement. The alternative hypothesis implicit in the ϕ-number treatment of the linear relation is deduced from the general theory.

1. The problem

Given n measurements (x_i, y_i) of two variates, each subject to error, and some theoretical basis for a structural relation governing the true points (X_i, Y_i), it is a fundamental problem in the physical and engineering sciences (Brown, 1955) to establish whether or not the structural relation is in accord with the evidence. One simple form of this problem is considered in the present paper; it is supposed that the conduct of the experiment has been judged reliable in the context of existing knowledge and that the data are so organized that the distribution of errors can be ascertained and adopted as a quantitative measure of reliability. The problem thus formulated was suggested by an experiment on coal breakage (Brown, 1947).

Previous work has been limited mainly to the linear relation and the main objective has been to estimate a best relation. When the method of maximum likelihood is used it is found that the best relation passes through the mean of the observations; the limits obtained for the coefficients defining the relation are then conditional on this requirement. Moreover, it is usual to find limits for the coefficients separately; this procedure destroys the structurality of the relation; it is essential to treat the relation *in toto*. Attempts to find relations amongst variates in the absence of independent information on the errors show that it is then necessary to assume that the relation exists, leading to doubtful inverted arguments; Moran (1956) has recently shown that the relation is then unidentifiable. It is well known that in the linear case the ratio of the error variances suffices for the estimation of a best relation; here the alternative hypotheses are not obvious and it is even doubtful if they can be formulated.

2. Linear relation

2·1. *ϕ-number*

If the data are normalized and the error* in x is distributed $N(0, 1)$ and that in y as $N(0, 1)$, then the perpendicular from an observation (x, y) on to a line

$$Y = \alpha_0 + \alpha_1 X \tag{2·1}$$

is distributed $N(0, 1)$. Since, as Lindley (1947) has pointed out, it is unnecessary to say anything about the true points (X, Y), part of the difficulty of the problem is removed. Putting

$$n\phi = \sum_i \frac{(-y_i + \alpha_0 + \alpha_1 x_i)^2}{(1 + \alpha_1^2)}, \tag{2·2}$$

* This conventional formulation is discussed further in § 3·1.

we know that $n\phi$ has a χ^2_n-distribution $\mathscr{L}(\phi)\,d\phi$. Then the numerical* statement,

$$\Pr(\phi \leqslant \phi_p) = \int_0^{\phi_p} \mathscr{L}(\phi)\,d\phi = p, \tag{2·3}$$

associates a probability p with a chosen ϕ_p. Notice that the differential element $d\phi$ is mixed, depending on both the observations and the coefficients of the line. Writing nS_{xx}, nS_{xy}, etc., for the sum of squares and products of the observations and taking the origin at the mean, we have

$$\phi = \frac{S_{yy} + \alpha_0^2 - 2\alpha_1 S_{xy} + \alpha_1^2 S_{xx}}{1 + \alpha_1^2}. \tag{2·4}$$

A confidence region for the coefficients (α_0, α_1) jointly can be written down immediately from equation (2·4). It is the non-central conic,

$$\alpha_1^2(S_{xx} - \phi_p) - 2\alpha_1 S_{xy} + \alpha_0^2 = (\phi_p - S_{yy}). \tag{2·5}$$

The coefficients (α_0, α_1) of all lines for which $\phi \leqslant \phi_p$ give points lying inside the conic, leading to the usual assertion that the coefficients lie in this region in a proportion p of cases in the long run. The region is bounded if the conic is an ellipse.†

2·2. *Acceptance conic*

Confidence theory is not developed further in this study of structural relations since it does not correspond to the habit of thought of a physicist who pictures the problem as ill-defined graphs in the space of true points (X, Y). He would regard the coefficients (α_0, α_1), even if predicted from a valid mathematical theory (Jeffreys, 1948, pp. 283–5), as uncertain and would wish to translate the ϕ-number (2·4) in terms of the probability statement (2·3) into a family of lines acceptable at probability level p (cf. § 3·1). Such a transformation of the probability distribution into true point space is easily effected by finding the envelope of lines (2·1) subject to the constraint (2·4) on the coefficients. This gives the *acceptance conic*,

$$\frac{(Y - a_1 X)^2}{(\phi_p - \phi_1)} - \frac{(a_1 Y + X)^2}{(\phi_2 - \phi_p)} = (1 + a_1^2), \tag{2·6}$$

where

$$a_1 = \frac{1}{2S_{xy}}\left[(S_{yy} - S_{xx}) + \{(S_{yy} - S_{xx})^2 + 4S_{xy}^2\}^{\frac{1}{2}}\right], \tag{2·7}$$

$$\phi_1 = S_{xx} - \frac{S_{xy}}{a_1}, \quad \phi_2 = S_{xx} + a_1 S_{xy}, \tag{2·8}$$

and it is assumed that $a_1 > 0$, $\phi_1 < \phi_2$. Equation (2·6) does not contain the unknown coefficients (α_0, α_1).

The acceptance conic could be regarded as a transformation into true point space of the confidence conic (2·5), and, as such, it could be given a confidence foundation. This view is supported by the disappearance of the unknown coefficients (α_0, α_1) in the process of obtaining the envelope. But the envelope is also to be regarded as the locus of all lines in true point space having $\phi = \phi_p$, and the locus is determined for the particular observations (x_i, y_i). Thus an interpretation analogous to fiducial theory appears feasible. In the author's

* In the sense of indifference to content as opposed to numerical measure; cf. the third regularity of § 3·1.

† The limits are the same as those found in § 2·2.

view, however, the fiducial method is not applicable since structural coefficients differ *in kind* from population parameters; in § 3·1 it is suggested that the acceptance conic should be interpreted in terms of new hypotheses of regularity. Meanwhile, it may be noted that, on account of the existence of a ϕ-number, the alternative interpretations yield the same limits for (α_0, α_1). It appears worth while to show that the acceptance conic has satisfactory properties. These are:

(i) the relation has been treated *in toto*,

(ii) the central conics for ϕ_p differing are confocal,

(iii) for $\phi_1 < \phi_p < \phi_2$, the conic is an hyperbola and all lines having $\phi < \phi_p$ lie inside the hyperbola, which therefore bounds all acceptable relations,

(iv) for $\phi_1 > \phi_p$, the conic is imaginary and it follows that there is no acceptable linear relation,

(v) for $\phi_p > \phi_2$, the conic is an ellipse and every line outside the ellipse has $\phi > \phi_p$; here it might be said that the line is indeterminate in direction, since the range of the data were not large enough in comparison with the errors; the ellipse is then the limiting boundary of the mean of the observations.

When the acceptance conic is an hyperbola, the limits for α_1 are given by the asymptotes

$$\text{limits}\,\alpha_1 = \frac{a_1\sqrt{(\phi_2-\phi_p)} \pm \sqrt{(\phi_p-\phi_1)}}{\sqrt{(\phi_2-\phi_p)} \mp a_1\sqrt{(\phi_p-\phi_1)}}, \tag{2·9}$$

and for the intercept I at $X = h$,

$$kI = a_1 h(\phi_2-\phi_1) \pm [(1+a_1^2)(\phi_2-\phi_p)(\phi_p-\phi_1)(k+h^2\overline{1+a_1^2})]^{\frac{1}{2}}, \tag{2·10}$$

where

$$k = (\phi_2-\phi_p) - a_1^2(\phi_p-\phi_1).$$

These intercepts do not belong to parallel lines. Notice that they depend on one probability level p. The axes of the hyperbola are

$$Y - a_1 X = 0, \quad a_1 Y + X = 0, \tag{2·11}$$

and these can be called respectively the 'best' and 'worst' lines through the mean of the observations. But the 'best' line is not here estimated in the usual sense.

Next, if r is the correlation coefficient of the observations,

$$r^2 = \frac{S_{xy}^2}{S_{xx}S_{yy}} = \left(1-\frac{\phi_1}{S_{xx}}\right)\left(1-\frac{\phi_1}{S_{yy}}\right) > \left(1-\frac{\phi_p}{S_{xx}}\right)\left(1-\frac{\phi_p}{S_{yy}}\right), \tag{2·12}$$

it being intuitive that $\phi_1 < S_{xx}, S_{yy}$. This might be regarded as a test for a unit population correlation ($\rho = 1$). In the example of § 2·3 below, $r^2 = 0{\cdot}94$ and $\left(1-\frac{\phi_p}{S_{xx}}\right)\left(1-\frac{\phi_p}{S_{yy}}\right) = 0{\cdot}90$ at $p = 0{\cdot}95$. Of course, (2·12) makes use of the errors of measurement and provides a test different from that derived from the bivariate normal distribution, wherein ρ cannot be unity.

Lastly, the method may be shown to be consistent. Thus the pairs of sums

$$S_x, S_X; \quad S_y, (\alpha_1 S_X + \alpha_0); \quad S_{xx}, (S_{XX}+1); \quad S_{xy}, \alpha_1 S_{XX}; \quad S_{yy}, (\alpha_1^2 S_{XX} + \alpha_0^2 + 1),$$

converge in probability to the same values, provided the errors are not correlated to the observations. Hence the pairs a_1, α_1; $\phi_1, 1$; $\phi_2, S_{XX}(1+\alpha_1^2)$, converge in probability to the same values. Also the hyperbola tends in the limit to the true line $Y = \alpha_1 X$ repeated.

The hyperbola is defined by the five statistics $\bar{x}$, $\bar{y}$, a_1, ϕ_1, ϕ_2, and these are unbiased and consistent in probability. These statistics give a sufficient expression of the data for the purpose of providing, through the ϕ-number, a test for the existence of a linear structural relation. Nevertheless, I consider that $n\phi$ has a χ^2_n-distribution. This question of degrees of freedom is discussed more fully in the extension of the problem to acceptance quadrics in m dimensions which it is hoped to publish in due course.

2·3. *Example*

The following data were prepared with the aid of a table of random normal deviates, the errors in x and in y being taken $N(0, 1)$:

x	1·8	4·1	5·8	7·5	9·3	10·6	13·4	14·7	18·9
y	6·9	12·5	20·0	15·7	24·9	23·4	30·2	35·6	39·1

We find

$$n = 9 \qquad \bar{x} = 9{\cdot}57 \qquad \bar{y} = 23{\cdot}14$$
$$nS_{xx} = 238 \qquad nS_{xy} = 451 \qquad nS_{yy} = 906$$
$$n\phi_1 = 11 \qquad n\phi_2 = 465 \qquad a_1 = 1{\cdot}98$$

Now χ^2_9 at $p = 0{\cdot}95$ gives $n\phi_p = 20$. Obviously there are acceptable straight lines, the acceptance hyperbola being

$$(Y-2X)^2 - 0{\cdot}02(2Y+X)^2 = 5,$$

referred to the mean as origin. The limits for the slope are

$$\alpha_1 = 1{\cdot}5 \text{ to } 2{\cdot}7,$$

and for α_0 (at $h = 0$)

$$\alpha_0 = \pm 2{\cdot}2.$$

The true line $(Y-\bar{y}) = 2(X-\bar{x}) - 1$, from which the data were derived, does not meet the acceptance hyperbola. (See Fig. 1.)

2·4. *Regression case* ($\sigma_x = 0$, $\sigma_y = 1$)

Here

$$n\phi = \sum_i (-y_i + \alpha_0 + \alpha_1 x_i)^2 \tag{2·13}$$

has a χ^2_n-distribution. Putting

$$a_1 = \frac{S_{xy}}{S_{xx}}, \quad nw = \frac{S_{xx}S_{yy} - S^2_{xy}}{S_{xx}} = \frac{1}{S_{xx}} \sum_{i<j} \left\{ \begin{vmatrix} 0 & 0 & 1 \\ x_i & y_i & 1 \\ x_j & y_j & 1 \end{vmatrix} \right\}^2, \tag{2·14}$$

the acceptance conic is

$$\frac{(Y-a_1X)^2}{(\phi_p - w)} - \frac{X^2}{S_{xx}} = 1, \tag{2·15}$$

which may be interpreted as before. Notice that in w the type term is zero if the pair of points (x_i, y_i), (x_j, y_j) lie on a line through the mean. The 'best' line is the regression line $Y - a_1X = 0$. The limits for α_1 are

$$\text{limits}\, \alpha_1 = a_1 \pm \left\{ \frac{\phi_p - w}{S_{xx}} \right\}^{\frac{1}{2}}, \tag{2·16}$$

and for the intercept I at $X = h$,

$$I = a_1 h \pm \left\{ (\phi_p - w)\left(1 + \frac{h^2}{S_{xx}}\right) \right\}^{\frac{1}{2}}. \tag{2·17}$$

Increasing S_{xx} narrows the range of both these limits, an outcome that corresponds to a usual expectation of a physicist.

When the error in the y-direction is based on a limited number of degrees of freedom, the same result holds with Snedecor's F-distribution replacing the χ^2-distribution of ϕ.

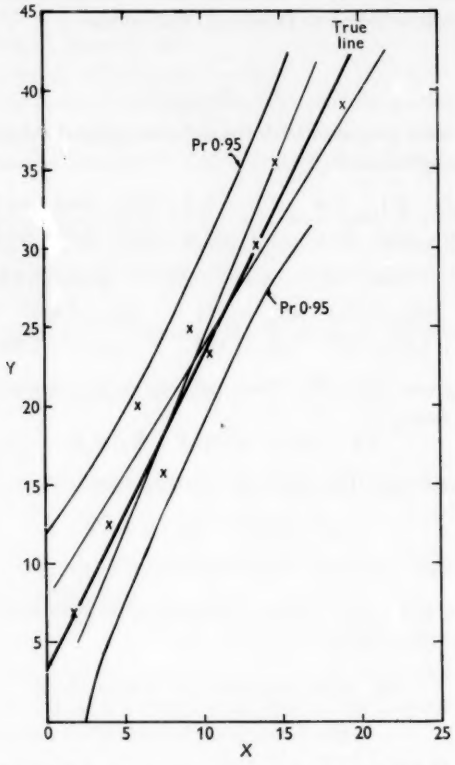


Fig. 1. Example of acceptance hyperbola derived from a ϕ-number.

2·5. *Further development*

Consider now a case when a ϕ-number having a simple probability distribution cannot be found. This arises when estimates, each on m degrees of freedom, are available for the error variances. Suppose the data to have been studentized. Then if p_i is the perpendicular from the ith observation on to the line (2·1),

$$p_i = \delta x_i \sin\theta - \delta y_i \cos\theta, \tag{2·18}$$

where δx_i, δy_i are the errors of the ith observation. Hence assuming there is no bias in the measurements, p_i is distributed $N(0, \sigma_x^2 \sin^2\theta + \sigma_y^2 \cos^2\theta)$, and thence $n\phi$ in (2·2) is distributed $\chi_n^2(\sigma_x^2 \sin^2\theta + \sigma_y^2 \cos^2\theta)$.

Now if $\delta x_s'$, $\delta y_s'$ are the errors in x, y from which the estimates $s_x^2 = s_y^2 = 1$ were derived,

$$p_s' = \delta x_s' \sin\theta - \delta y_s' \cos\theta \tag{2·19}$$

is normally distributed, and therefore $\sum_1^m p_s'^2$ is distributed as chi-squared. But

$$E\left(\sum_1^m p_s'^2\right) = m(\sigma_x^2 \sin^2\theta + \sigma_y^2 \cos^2\theta), \tag{2·20}$$

provided

$$E(\delta x_s' \delta y_s') = 0, \tag{2·21}$$

which is usually the case. Since the observations (x_i, y_i) are studentized,

$$\sum_1^m \delta x_s'^2 = \sum_1^m \delta y_s'^2 = m. \tag{2·22}$$

Put

$$\sum_1^m \delta x_s' \delta y_s' = mh. \tag{2·23}$$

Then

$$\sum_1^m (\delta x_s' \sin\theta - \delta y_s' \cos\theta)^2 = m(1 - 2h \sin\theta \cos\theta) \tag{2·24}$$

has a $\chi_m^2(\sigma_x^2 \sin^2\theta + \sigma_y^2 \cos^2\theta)$ distribution. Thus

$$f = \frac{\phi}{1 - \dfrac{2h\alpha_1}{1+\alpha_1^2}} \tag{2·25}$$

has an $F(n, m)$ distribution.

This result is of little value since the quantity h would not usually be known. h has no particular physical significance under usual conditions of experimentation.

Further complications arise when, (*a*) it is not assumed that the measurements are free from bias, (*b*) the degrees of freedom with which s_x^2, s_y^2 are determined are not equal.

In such cases, the methods proposed here, coupled with approximate distribution theory, would appear to lead to approximate tests. Perhaps the lack of an exact statistical test is not unexpected, since the organization of additional information in the form of errors of measurement requires a combination of scientific and probabilistic languages that merits deep consideration.*

3. CURVILINEAR RELATIONS

3·1. *The three hypotheses of regularity*

The simplicity of the linear problem is due to the existence of a ϕ-number, leading to a confidence theory or to an acceptance conic. But any satisfactory approach to structural relations, as they are understood by physicists and engineers, must be capable of leading to curvilinear systems. A more general treatment is necessary even in the case of the linear relation, for it will have been noted that the alternative hypothesis of a structural, but non-linear, relation is not encompassed in the derivation from the ϕ-number. A new theory will now be given for the general relation

$$Y = f(X, \alpha_j) \quad (j = 0, 1, \ldots, s-1). \tag{3·1}$$

For a physicist, there is a pattern of relationships underlying his scientific thought. In other words, the disposable parameters α_j are not uniquely determined; their values are

* Kendall (1951, 1952) and others have approached tentatively this question in their discussion of the different types of variate that are encountered in practice. See also § 3·1.

uncertain.* Moreover, the functional forms $f(X_1\alpha_j)$ can translate into each other, usually, however, within a limited class of functions such as wave or power laws.

The meaning of (3·1) can be clarified by enunciating three hypotheses of regularity:

(i) *Regularity in the succession of observational data.* This hypothesis is implicit in confidence theory in so far as it is considered to yield results valid in the long run. Adopting a much simplified argument, if X is a true value, x an observation, then

$$\epsilon = (X - x) \tag{3·2}$$

defines an error ϵ. In confidence theory X is regarded as knowable absolutely and the relation $d\epsilon = -dx$ puts all the error on to the observational data. Then the probability distribution $p(\epsilon)\,d\epsilon$ is regarded as established by repeated observation.

(ii) *Regularity in the totality of coexisting patterns of structural relations.* If the observational data are regarded as known and fixed, i.e. a subset of all possible data is agreed to be chosen, then the relation $d\epsilon = dX$ yields what may be called a *structural theory*, the development of which has been the objective of this paper. As an example, consider the class of patterns of structural relations that are linear. This class is abstracted from a wider class on the supposition that the disturbances of linearity are many and small. Such abstraction may be considered valid if account be taken of the probability distribution $p(\epsilon)\,d\epsilon$, found now from the central limit theorem and not by repeated observation.

(iii) *Regularity in the recurrent alternations between underlying patterns and observational data.* Alternation of theory and experiment is a recognized element in scientific work. It reconciles the opposition of the first two regularities, taking into account that patterns in their totality and the observational data in all their possible actualizations are both unknowable, though for different reasons. Recurrent alternation is the true source of the error ϵ and its probability distribution, which strictly apply to neither of the first two regularities. If ϵ is distributed $N(\mu, \sigma)$, then the relation $d\epsilon = d\mu$ would appear to lead to fiducial theory; the author at present considers that the fiducial theory of the location parameter μ is often a structural theory in the sense defined here, whereas that of the scale parameter σ may be distinct theory, neither confidence nor structural.

It will be noticed that the first regularity cannot in fact yield results 'in the long run', although the introduction of randomization procedures can in part overcome this difficulty. Thus it is not straightforward to accept it without appealing tacitly to an intuitive appreciation of the determination of occasions in the framework of space-time, with a consequent limited notion of repetition. A generalized notion of repetition is needed to accommodate all three regularities and this requires an ampler framework, such as has been proposed, for quite independent reasons, by physical scientists.† In such a framework all three regularities must be accepted as determining what can be known; they gain intelligibility through the

* Jeffreys (1948, p. 60) identifies a true value with the mean of the probability distribution; but later (pp. 268–76), in discussing systematic errors, he distinguishes the true value from the location parameter. The distinction is not trivial. See also the 'Principle of Limited Variety' in C. D. Broad's presidential address to the Aristotelian Society (1927–8).

† For example, Kaluza (1921), Caldirola (1942), Bennett, Brown & Thring (1949), and others have explored the scope offered by a five-dimensional framework for the representation of physical events. Bennett has emphasized especially the anti-time-like character of a fifth dimension. Podalanski (1950) and Bennett (1953) have proposed a six-dimensional framework. Bennett's precise characterization of the two extra dimensions gives a clear basis for the three regularities cited here. Of course, all three regularities determine every occasion.

relativistic notion of repetition due to Bennett (1957), namely, that the moment of observation, with its patterns of potential possibilities, is multi-valued in its actualizations.

It follows from the second regularity that in a structural theory the relationship should be treated *in toto* and should embody the alternative relationships within the class abstracted from the totality of all possible patterns.

The question then arises as to how an independent and relevant assessment of the probability distribution can be made. This might be derived theoretically or experimentally. The former rests on the central limit theorem. In the latter case, consider the coal experiment that led to this investigation, where the errors were obtained from measurements of sub-populations clearly defined by an attribute meaningful in respect to the measurements x, y and the structural relation under consideration. The attribute was determined by the third regularity. In the absence of such an attribute it is questionable whether an adequately reliable experiment can be done; the adaptation necessary in Wald's (1940) treatment, in which the observations (x_i, y_i) are divided into two sets, is not obvious and perhaps not appropriate.

3·2. *Structural theory*

Suppose that n measurements (x_i, y_i) and independent information relating to the distributions $p(\epsilon)\,d\epsilon$, $q(\theta)\,d\theta$ of the errors (ϵ, θ) associated with (x, y) respectively are given, these being supposed to be free from bias. When X is given in (3·1), Y can be calculated. Let $X_i (i = 1, \ldots, n)$ give the location of the true points corresponding to the observations (x_i, y_i). Thence the $2n$ errors are

$$\epsilon_i = (X_i - x_i), \quad \theta = (f_i - y_i), \tag{3·3}$$

where f_i is evaluated at X_i.

Take $s = n$. Regarding the coefficients α_j as lying in an n-dimensional coefficient space $(\alpha_0, \alpha_1, \ldots, \alpha_{n-1})$, the true coefficients may be thought of as a point. Then the errors θ are given by variation of points in the coefficient space, i.e. by variation of α_j.

The joint probability distribution of the $2n$ errors ϵ_i, θ_i is

$$p(d\epsilon_1, d\epsilon_2, \ldots, d\epsilon_n, d\theta, \ldots, d\theta_n) = \prod_{i=1}^{n} p(\epsilon_i)\, q(\theta_i)\, d\epsilon_i d\theta_i. \tag{3·4}$$

Transform to $2n$ new variables X_i, α_j. The Jacobian is easily seen to be the $n \times n$ determinant

$$\left\| \frac{\partial f_i}{\partial \alpha_j} \right\|,$$

and the following equation

$$p(dX_1 \ldots dX_n, d\alpha_0 \ldots d\alpha_{n-1}) = \left\| \frac{\partial f_i}{\partial \alpha_j} \right\| \prod_{i=1}^{n} \prod_{j=0}^{n-1} p(X_i - x_i)\, q(f_i - y_i)\, dX_i\, d\alpha_j \tag{3·5}$$

is obtained.

This equation, which is basic for what follows, is interpreted as a probability density in the $2n$-fold of the n true values X_i and the n-dimensional coefficient space α_j. The X_i locate the true points on the curve and the α_j locate the curve having the functional form given in (3·1).

Suppose now that the structural relation is a simple additive function of the disposable coefficients α_j, i.e.

$$f(X, \alpha_j) = \sum_{j=0}^{n-1} \alpha_j \phi_j(X), \tag{3·6}$$

where $\phi_j(X)$ are n independent functions of X only. Then

$$p\Big(\prod_1^n dX_i \prod_0^{n-1} d\alpha_j\Big) = \| \phi_j(X_i) \| \prod_{i=1}^n \prod_{j=0}^{n-1} p(X_i - x_i)\, q(f_i - y_i)\, dX_i d\alpha_j. \tag{3·7}$$

It may be noted that the n functions $\phi_j(X)$ contain in themselves all the alternative hypotheses, as is necessary in a structural theory.

To proceed further from equation (3·5) or (3·7) it is necessary to introduce an hypothesis for the true points X_i. Strictly, this hypothesis should be related to the way in which the experiment is carried out. Thus a physicist would endeavour so to space his data along the curve as to give it good definition. However, it does not appear to be convenient to introduce such information into a general treatment, especially bearing in mind that it is uncertain whether the physicist has achieved (or even, can achieve) his objective.

So long as there are n disposable parameters α_j, the most probable positions of X_i are the observations x_i themselves. But given an hypothesis that in the coefficient space of the α's there is a region such that some of the α's are zero, then the likelihood may be maximized with respect to the X_i, leaving a joint probability distribution for the remaining α's.

A third method is to suppose that each true point can be located anywhere on the curve and to form the n-fold integral of equation (3·5) or (3·7) over the range of possible variation of each of the X_i. This method is adopted here. It has the merit of stressing the structural nature of the relation (3·1).

3·3. *Two observations and normally distributed errors*

When the errors are normally distributed and adjusting the scale of the observations so that both standard deviations are unity,

$$p(\epsilon)\, d\epsilon = \frac{1}{\sqrt{(2\pi)}} e^{-\frac{1}{2}\epsilon^2} d\epsilon, \quad q(\theta)\, d\theta = \frac{1}{\sqrt{(2\pi)}} e^{-\frac{1}{2}\theta^2} d\theta. \tag{3·8}$$

Here bias has been removed by taking zero means. Consider the linear relation

$$Y = \alpha_0 + \alpha_1 X. \tag{3·9}$$

With two observations, equation (3·7) becomes

$$\begin{aligned} p(d\alpha_0 d\alpha_1) = \int_{X_1=-\infty}^{\infty} \int_{X_2=-\infty}^{\infty} & \frac{1}{2\pi} |\,(X_1 - X_2)\,| \\ & \times \exp\left[-\tfrac{1}{2}(X_1 - x_1)^2 - \tfrac{1}{2}(X_2 - x_2)^2 - \tfrac{1}{2}(-y_1 + \alpha_0 + \alpha_1 X_1)^2\right. \\ & \left. -\tfrac{1}{2}(-y_2 + \alpha_0 + \alpha_1 X_2)^2\right] dX_1 dX_2 d\alpha_0 d\alpha_1, \end{aligned} \tag{3·10}$$

whence

$$\begin{aligned} p(d\alpha_0 d\alpha_1) = & \left| \frac{\alpha_1(y_1 - y_2) + (x_1 - x_2)}{2\pi(1+\alpha_1^2)^2} \right| \\ & \times \exp\left[-\frac{1}{(1+\alpha_1^2)} \left\{ \left(\alpha_0 - \frac{y_1+y_2}{2} + \alpha_1 \frac{x_1+x_2}{2} \right)^2 + \tfrac{1}{4}(\alpha_1 \overline{x_2 - x_1} - \overline{y_2 - y_1})^2 \right\} \right] d\alpha_0 d\alpha_1. \end{aligned} \tag{3·11}$$

As might be expected, the probability distribution of α_1 is independent of α_0, whereas the distribution of α_0 contains α_1.

Transform to independent variates u, v by writing $\tan\theta$ for α_1 and

$$\left.\begin{aligned}\sqrt{2}\,u &= (x_2-x_1)\sin\theta-(y_2-y_1)\cos\theta,\\ \frac{v}{\sqrt{2}} &= (\alpha_0-\tfrac{1}{2}\overline{y_1+y_2})\cos\theta+\tfrac{1}{2}(x_1+x_2)\sin\theta.\end{aligned}\right\} \tag{3·12}$$

$v/\sqrt{2}$ is the length of the perpendicular from the midpoint of the join of the two observations on to the line (3·9) and $u\sqrt{2}$ is the length of the projection, perpendicular to the line (3·9), of the join of the two observations. Then u, v are each normally distributed with zero mean and unit variance. The modal values of u, v correspond to the line passing through the observations; for this line

$$\hat{\alpha}_0 = \left(\frac{x_2y_1-x_1y_2}{x_2-x_1}\right), \quad \tan\hat{\theta} = \left(\frac{y_2-y_1}{x_2-x_1}\right). \tag{3·13}$$

3·4. *The ϕ-number for the linear relation*

Given $2n$ observations and normally distributed errors, write

$$\left.\begin{aligned}\sqrt{2}\,u_r &= (x_{2r}-x_{2r-1})\sin\theta_r-(y_{2r}-y_{2r-1})\cos\theta_r,\\ v_r/\sqrt{2} &= (\alpha_{0r}-\tfrac{1}{2}\overline{y_{2r}+y_{2r-1}})\cos\theta_r+\tfrac{1}{2}(x_{2r}+x_{2r-1})\sin\theta_r,\end{aligned}\right\} \tag{3·14}$$

so that the observations $(x_{2r-1}, y_{2r-1})\,(x_{2r}, y_{2r})$ are associated with the line

$$Y = \alpha_{0r} + X\tan\theta_r. \tag{3·15}$$

Then each of u_r, v_r is distributed $N(0, 1)$ and they are evidently independent. Thus

$$\begin{aligned}2n\phi &= \sum_{r=1}^{n}(u_r^2+v_r^2)\\ &= \sum_{r=1}^{n}[\{(-y_{2r}+\alpha_{0r})\cos\theta_r+x_{2r}\sin\theta_r\}^2+\{(-y_{2r-1}+\alpha_{0r})\cos\theta_r+x_{2r-1}\sin\theta_r\}^2]\end{aligned} \tag{3·16}$$

is distributed as χ^2_{2n}. Now for the linear relation

$$\alpha_{0r} = \alpha_0, \quad \theta_r = \theta, \tag{3·17}$$

and equation (3·15) becomes exact. Also the χ^2_{2n}-distribution yields a likelihood $\mathscr{L}$ in ϕ or numerical space: hence a structural theory for the use of ϕ-numbers in § 2. It is interesting to note that the particular way in which this result is derived expresses the alternative hypothesis in the form that one or more pairs of observations do not lie on a straight line through the mean. It follows that when in § 2 there is no real locus of equally probable lines, alternative hypotheses of the form indicated in this paragraph must be considered. More fundamentally, alternatives deriving from the pattern of structural relations, e.g. that there is a parabolic or higher degree relation, may be considered. For the latter a ϕ-number has not yet been found.

The validity of the linear hypothesis (3·17) may also be tested by examining whether there are some (α_0, θ) for which u_r, v_r are normal deviates. To this end, statistics such as the mean, sum of squares and $k_3k_2^{-\frac{3}{2}}$, where the k's are cumulants, might be computed and compared with their known distributions. Such tests are only approximate in virtue of equation (3·15).

3·5. *The general case*

For normally distributed errors, the general case remains unsolved. For example a linear relation as a limiting case of a parabola cannot yet be obtained. The difficulties are considerable. Those to be expected in any problem of curve fitting become evident, with the further complication that the slope of the curve in the neighbourhood of each observation enters into the compounded error variance, so that the equations are non-linear. An indication of an approximate method—specifically applied to a parabolic relation—is given in an appendix. It is hoped to discuss elsewhere further developments from equation (3·5) having in mind that the Jacobian contains in itself the whole scheme of alternative hypotheses that can be introduced into a discussion of any class of bivariate structural relationship.

4. Conclusion

Whether or not the physicist or engineer needs to use statistical methods of inference is perhaps beside the point. His procedure consists of a series of steps. First, an endeavour is made to isolate certain phenomena. Secondly, an attempt is made to reduce them to measurements by conventional rules, these being generally accepted by scientists. Thirdly, the results of reliable experiments are compared with (*a*) general background knowledge and (*b*) present theoretical (structural) conceptions. Then a fresh experiment is designed. The interesting question is whether an investigation of the type undertaken here can show the strength and weakness of the several steps, adopted intuitively or unconsciously, by the practising scientist. For this reason, further investigations on the present lines appear to be necessary.

The author is indebted to Mr F. Fereday, Mr S. R. Broadbent and Mr W. D. Ray for several useful discussions during the course of the investigations and to Mr J. G. Bennett for guidance in formulating his ideas in the form of the three regularities given in § 3·1.

REFERENCES

Bennett, J. G. (1953). *Proc. 11th Int. Congr. Philosophy*, **6**, 102.
Bennett, J. G. (1957). *The Dramatic Universe*. **1**. *The Foundations of Natural Philosophy*. London: Hodder and Stoughton.
Bennett, J. G., Brown, R. L. & Thring, M. W. (1949). *Proc. Roy. Soc.* A, **198**, 39.
Broad, C. D. (1927–8). The principles of problematic induction. *Proc. Arist. Soc.* p. 1.
Brown, R. L. (1955). Essence of design of experiments. *Appl. Mech. Rev.*, October.
Brown, R. L. (1947). The brittle fracture of pre-cracked solids. *Research*, **1**.
Caldirola, P. (1942). *Nuovo. Cim.* **19**, 25.
Jeffreys, H. (1948). *Theory of Probability*, 2nd ed. Oxford University Press.
Kaluza, A. (1921). *S.B. preuss. Akad. Wiss.* p. 968.
Kendall, M. G. (1951, 1952). Regression, structure and functional relationships. *Biometrika*, **38**, 11, and **39**, 96.
Lindley, D. V. (1947). Regression lines and the linear functional relationship. *Suppl. J. R. Statist. Soc.* **9**, 218.
Moran, P. A. P. (1956). A test of significance for an unidentifiable relation. *J. R. Statist. Soc.* B, **18**, 61.
Podalanski, J. (1950). *Proc. Roy. Soc.* A, **201**, 234.
Wald, A. (1940). The fitting of straight lines if both variables are subject to error. *Ann. Math. Statist.* **11**, 284.

APPENDIX

Approximate treatment of the curvilinear relation by relaxation of constraints

The exact solution given for two observations and normally distributed errors can be used to treat approximately the general case. First we relax the constraints on the observations imposed by the structural relation. We have to assume that the observations are fairly evenly spaced along the curve and that there are no 'repeat' observations. Repeat observations are equivalent to a reduction in the degree of the curve and would require, in any theory, special treatment. We use the approximation (3·15) to find the intercepts k_s of the curve on an equidistant grid of ordinates with abscissae $(sh+l)$. The grid is chosen so that the abscissa $(sh+l)$ is close to the observed value x_s, there being $2n$ lines on the grid. Then choosing the origin so that l is zero:

$$\left.\begin{aligned} k_{2r} &= \alpha_{0r} + (2rh)\tan\theta_r, \\ k_{2r-1} &= \alpha_{0r} + (\overline{2r-1}h)\tan\theta_r, \end{aligned}\right\} \tag{A 1}$$

and thence

$$\left.\begin{aligned} \alpha_{0r} &= (2rk_{2r-1} - \overline{2r-1}k_{2r}), \\ \tan\theta_r &= \frac{1}{h}(k_{2r} - k_{2r-1}). \end{aligned}\right\} \tag{A 2}$$

In this case

$$w_r^2 = (u_r^2 + v_r^2) = \frac{\{-y_{2r} + k_{2r} + \overline{k_{2r} - k_{2r-1}}\,\overline{x_{2r/h} - 2r}\}^2}{1 + (k_{2r} - k_{2r-1})/h^2} + \frac{\{-y_{2r-1} + k_{2r-1} + \overline{k_{2r} - k_{2r-1}}\,\overline{x_{(2r-1)/h} - 2r + 1}\}^2}{1 + (k_{2r} - k_{2r-1})/h^2}. \tag{A 3}$$

The constraints on k_s arising from a structural relation are now easily introduced. For example, a polynomial of degree m gives

$$k_s = \sum_{i=0}^{m} \alpha_i (sh)^i, \tag{A 4}$$

there being $(m+1)$ unknown parameters α_i. Since $2n\phi$, or $\sum_1^n w_r^2$, has a χ^2-distribution, minimum χ^2 estimators (also maximum-likelihood estimators) may then be found for the α_i. The minimum value of $2n\phi$ has then an approximate $\chi^2_{(2n-m-1)}$ distribution. The degree of the polynomial necessary for representation of the data can thus be found. Owing to the form of equation (A 3), it is necessary in practice to adopt an iterative method for finding the parameters α_i.

Parabolic relation

The approximate treatment of the preceding sub-section yields a statistical test for establishing the degree m of a polynomial (A 4) that adequately represents a structural relation between two variates, each subject to normal errors, the alternative hypothesis being that a polynomial of degree $(m+1)$ or greater is not necessary. Consider the parabolic relation

$$Y = \alpha_0 + \alpha_1 X + \alpha_2 X^2. \tag{A 5}$$

When the points on the grid (sh, k_s) lie on this parabola, we have from (A 3)

$$\left.\begin{aligned} \alpha_{0r} &= \alpha_0 - \alpha_2 h^2\, 2r(2r-1), \\ \tan\theta_r &= \alpha_1 + \alpha_2 h(4r-1). \end{aligned}\right\} \tag{A 6}$$

Then the equations determining α_0, α_1, α_2 for minimum $2n\phi$ are

$$\left.\begin{aligned} &\sum_{r=1}^{n} a_r \cos\theta_r = 0, \\ &\sum_{r=1}^{n} c_r \cos^2\theta_r = 0, \\ &\sum_{r=1}^{n} 2rc_r \cos^2\theta_r - hr\overline{2r-1}\, a_r \cos\theta_r = 0, \end{aligned}\right\} \tag{A 7}$$

where

$$\left.\begin{aligned} a_r &= (2\alpha_{0r} - \overline{y_{2r} + y_{2r-1}})\cos\theta_r + (x_{2r} + x_{2r-1})\sin\theta_r, \\ c_r &= \tfrac{1}{2}\sin 2\theta_r[2(x_{2r}^2 + x_{2r-1}^2) - (y_{2r} - y_{2r-1})^2 - (2\alpha_{0r} - y_{2r} - y_{2r-1})^2] \\ &\quad + 2\cos 2\theta_r[\alpha_{0r}(x_{2r} + x_{2r-1}) - (x_{2r}y_{2r} + x_{2r-1}y_{2r-1})]. \end{aligned}\right\} \tag{A 8}$$

Let $\alpha_0^1, \alpha_1^1, \alpha_2^1$ be an approximate solution of equation (A 7) and $\alpha_{0r}^1, \theta_r^1, a_r^1, c_r^1$ the corresponding quantities from equations (A 6, 8). Then expanding (A 7) by Taylor's theorem, we find the approximate linear simultaneous equations for $\Delta\alpha_0, \Delta\alpha_1, \Delta\alpha_2$, where $\Delta\alpha_0 = (\alpha_0 - \alpha_0^1)$, etc., to be

$$\Sigma 2\cos^2\theta_r^1\Delta\alpha_0 + \Sigma b_r^1\cos^2\theta_r^1\Delta\alpha_1 + \Sigma\{b_r^1\overline{4r-1} - h\,4r\overline{2r-1}\}\times\cos^2\theta_r^1 h\Delta\alpha_2 + \Sigma a_r^1\cos\theta_r^1 = 0,$$

$$\Sigma 2b_r^1\cos^2\theta_r^1\Delta\alpha_0 + \Sigma e_r^1\cos^2\theta_r^1\Delta\alpha_1 + \Sigma\{e_r^1\overline{4r-1} - hb_r^1\,4r\,\overline{2r-1}\}\times\cos^2\theta_r^1 h\Delta\alpha_2 + \Sigma c_r^1\cos^2\theta_r^1 = 0,$$

$$\begin{aligned}\Sigma 2r\cos^2\theta_r^1(2b_r^1 - h\overline{2r-1})\,\Delta\alpha_0 + \Sigma r\cos^2\theta_r^1\{2e_r^1 - h(2r-1)\,b_r^1\}\,\Delta\alpha_1 \\ + \Sigma r\cos^2\theta_r^1\{2\,\overline{4r-1}\,e_r^1 - hb_r^1\overline{2r-1}\,1\overline{2r-1} + h^2 4r\,\overline{2r-1}^2\}\,h\Delta\alpha_2 \\ + \Sigma(2c_r^1\cos\theta_r^1 - h\overline{2r-1}\,a_r^1)\,r\cos\theta_r^1 = 0,\end{aligned} \tag{A 9}$$

where

$$\begin{aligned}b_r &= (x_{2r}+x_{2r-1})\cos 2\theta_r - (2\alpha_{0r} - y_{2r} - y_{2r-1})\sin 2\theta_r,\\ e_r &= \{2(x_{2r}^2 + x_{2r-1}^2) - (y_{2r} - y_{2r-1})^2 - (2\alpha_{0r} - y_{2r} - y_{2r-1})^2\}\\ &\quad\times\left\{\cos 2\theta_r\cos^2\theta_r - \frac{\sin^2 2\theta_r}{2}\right\} - \{2\sin 2\theta_r(2\cos^2\theta_r + \cos\theta_r)\}\\ &\quad\times\{\alpha_{0r}(x_{2r}+x_{2r-1}) - (x_{2r}y_{2r} + x_{2r-1}y_{2r-1})\}.\end{aligned} \tag{A 10}$$

Using equations (A 9) an iterative method may be employed to find the values of $\alpha_0, \alpha_1, \alpha_2$ that minimize $(2n\phi)$. The first approximation may be found easily from the equation of the parabola through three likely points on the curve.

Note added by Author in proof

Professor M. G. Kendall has kindly drawn my attention to a paper in which Working and Hotelling (1929)* calculate the variances of the slope and intercept of a linear regression and show that these define an hyperbolic region within which the 'true' line lies with an assigned probability. Their hyperbola is similar to that given in equation (2·15) above, the denominator of the first term $(Y - a_1X)^2$ being (χ_n^2) instead of $(\theta_p - w)$. Working and Hotelling ascribe two degrees of freedom to χ^2, due to a (presumed) independent variation of the slope and intercept of the line: I have been unable to follow their reasoning on this point. In my case, $n\theta_p$ has a χ_n^2-distribution, but I have called for independent knowledge of the variance in the error of the y-measurement.

The two papers deal with different questions; in particular, I was mainly concerned with the case when both variates are in error. But the duality of the confidence ellipse and the acceptanee hyperbola noted in my section 2·2 is anticipated in the Working and Hotelling paper.

* H. Working & H. Hotelling (1929). Applications of the theory of error to the interpretation of trends. *J. Amer. Statist. Ass.* Suppl. **24**, 73.

AN ANALYSIS OF PAIRED COMPARISON DESIGNS WITH INCOMPLETE REPETITIONS†

By JOHN W. WILKINSON

University of North Carolina

1. Introduction

Experiments involving paired comparisons have mainly concerned the situation where each judge compares all possible pairs of treatments or objects. In certain types of experiments, this may require an excessive number of comparisons to be made by any observer. To overcome this handicap, Bose (1956) and Kendall (1955) constructed certain designs, symmetrical with respect to objects and judges, which do not require each judge to compare all possible pairs of objects. However, neither Bose nor Kendall proposed any procedure of analysis in their respective papers.

The purpose of this paper is to consider the problem of analysis of the Bose–Kendall paired comparison designs. This analysis is carried out in the tradition of the fundamental Bradley–Terry (1952) paper concerning the situation where all possible pairs are compared by each judge. Using the Bradley–Terry mathematical model, likelihood ratio tests are constructed in detail for certain classes of hypotheses, and are stated for some additional situations of interest. To exemplify the test procedures, the proposed analysis is applied to an experiment involving pairwise comparison of handwriting specimens.

2. Definition of paired comparison designs

A paired comparison design, as defined by Bose, is as follows:

Suppose we have t objects which we desire to compare according to a certain characteristic, and we have v judges available to perform the comparisons. The procedure of comparison will be as follows:

(i) Each judge compares r pairs of objects ($1 < r \leqslant \frac{1}{2}t(t-1)$), and, for each pair compared, expresses his preference for one or the other object of the pair. (It may be desirable to allow the judge to express no preference with respect to either of the objects forming the pair. However, this possibility will not be entertained in this paper.)

(ii) The pairs compared by any judge are all different.

(iii) Among the r pairs compared by each judge, each object appears equally often, say α times.

(iv) Each pair is compared by k judges ($1 < k \leqslant v$).

(v) Given any two judges, there are exactly λ pairs which are compared by both judges.

Bose constructed three series of paired comparison designs by making certain correspondences with known balanced incomplete block designs. One of these series produced the same designs as those constructed by Kendall.

† This research was jointly supported by the United States Air Force through the Air Force Office of Scientific Research of the Air Research and Development Command, and the National Research Council of Canada.

From the relationships among the parameters of balanced incomplete block designs, and from the above definition of paired comparison designs, there exist the following relationships among the parameters of a paired comparison design:

$$r = \tfrac{1}{2}t\alpha, \quad \lambda(v-1) = r(k-1), \quad v\alpha = k(t-1), \quad t\alpha \geqslant 2k, \quad \lambda \leqslant \tfrac{1}{2}\alpha(\alpha+1) \leqslant r. \tag{2·1}$$

3. Experimental procedure and notation

Suppose we have available v judges, and we require each of these judges to compare r pairs of the $\frac{1}{2}t(t-1)$ possible pairs of a set of t objects $(a_1, a_2, \ldots, a_t)$, $(1 < r \leqslant \frac{1}{2}t(t-1))$. The r pairs to be compared by each judge will be specified by the field plan of the appropriate paired comparison design employed. Presently available designs are given by Bose (1956) and by Wilkinson (1956). We shall refer to one set of r pairs compared by a specific judge as an incomplete repetition or a complete repetition, depending on whether or not $r < \frac{1}{2}t(t-1)$ or $r = \frac{1}{2}t(t-1)$. When $r = \frac{1}{2}t(t-1)$, the situation becomes that which is considered in detail by Bradley & Terry (1952), Bradley (1954*a*, *b*, 1955) and Terry, Bradley & Davis (1952).

For each pair of objects compared by judge u $(u = 1, \ldots, v)$, the judge expresses a preference for one object or the other, assigning the rank one to the object judged superior, and the rank two to the object judged inferior.

Let r_{iju} denote the rank assigned by judge u to the ith object when it is compared with the jth object. Let $(a_i \rightarrow a_j \,|\, u)$ denote the preference indicated by judge u for object a_i over object a_j $(i \neq j,\ i, j = 1, \ldots, t;\ u = 1, \ldots, v)$. Then

$$r_{iju} = \begin{cases} 1, & \text{if} \quad (a_i \rightarrow a_j \,|\, u), \\ 2, & \text{if} \quad (a_i \leftarrow a_j \,|\, u), \end{cases} \tag{3·1}$$

if objects a_i and a_j are compared by judge u. From (3·1) it is readily observed that

$$r_{iju} + r_{jiu} = 3, \tag{3·2}$$

when objects a_i and a_j are compared by judge u. To handle the complication created by the fact that each judge does not compare all possible pairs of objects, we shall define

$$n_{iju} = \begin{cases} 1, & \text{if} \quad a_i \text{ and } a_j \text{ are compared by judge } u, \\ 0, & \text{if} \quad a_i \text{ and } a_j \text{ are not compared by judge } u, \end{cases} \tag{3·3}$$

$(i \neq j,\ i, j = 1, \ldots, t;\ u = 1, \ldots, v)$, and we shall conventionally take

$$n_{iiu} = 0 \quad (i = 1, \ldots, t;\ u = 1, \ldots, v).$$

Hence, for each judge u we can define an *incidence matrix* $N_u = (n_{iju})$ which is symmetric, has 1's or 0's for elements, and is specified by the field plan of the paired comparison design used. These matrices N_u satisfy the conditions

$$\left.\begin{aligned} &(a) \quad N_u E = \alpha E \quad (u = 1, \ldots, v), \\ &(b) \quad \sum_{u=0}^{v} N_u = kE, \\ &(c) \quad \operatorname{tr} N_u N_{u'} = 2\lambda \quad (u, u' = 1, \ldots, v;\ u \neq u'), \end{aligned}\right\} \tag{3·4}$$

where E is a matrix of order t, each of whose elements is unity, and where $N_0 = kI$, I being the identity matrix of order t. In fact, these conditions are the necessary and sufficient conditions for the matrices N_u to specify a paired comparison design as defined in § 2.

Conditions (*a*), (*b*), (*c*) are directly necessary for the conditions (iii), (iv), (v) of § 2. Condition (ii) is satisfied since N_u has only 0 and 1 as elements. Since $2r$ is the number of unities in N_u, the constancy of r follows from (*a*), which implies the number of unities in each row of N_u to be α. Hence $2r = t\alpha$.

To simplify notation, we will introduce the following conventions: $\sum_i$, $\sum_j$, $\sum_u$, $\sum_m$ and $\prod_i$, $\prod_j$, $\prod_u$, $\prod_m$ will indicate, respectively, single sums and products with respect to the depicted quantity over its range, where $i = 1, \ldots, t$; $j = 1, \ldots, t$; $u = 1, \ldots, v$; $m = 1, \ldots, g$. $\sum_j'$ will denote $\sum_{j=1, j\neq i}^{t}$ for some i, where this i will be clear from the context. $\sum_{i<j}$ and $\prod_{i<j}$ will indicate, respectively, double sums and products, $i = 1, \ldots, j-1$; $j = 2, \ldots, t$. $\sum_{i\neq j}$ and $\prod_{i\neq j}$ will indicate, respectively, $\sum_{\substack{i=1 \\ i\neq j}}^{t} \sum_{j=1}^{t}$ and $\prod_{\substack{i=1 \\ i\neq j}}^{t} \prod_{j=1}^{t}$. Any departures from these conventions will be specified when the departure is incurred. In addition, log and ln will be used to denote common and natural logarithms, respectively.

4. MATHEMATICAL MODEL (BRADLEY–TERRY MODEL)

Suppose there exist numbers $\pi_{1u}, \ldots, \pi_{tu}$ corresponding to judge u ($u = 1, \ldots, v$), where

$$\left.\begin{aligned} \pi_{iu} &\geqslant 0 \quad (i = 1, \ldots, t;\ u = 1, \ldots, v), \\ \sum_i \pi_{iu} &= 1 \quad (u = 1, \ldots, v), \end{aligned}\right\} \tag{4·1}$$

and such that

$$\left.\begin{aligned} P(a_i \rightarrow a_j \mid u) &= \frac{\pi_{iu}}{\pi_{iu} + \pi_{ju}}, \\ P(a_i \leftarrow a_j \mid u) &= \frac{\pi_{ju}}{\pi_{iu} + \pi_{ju}}, \end{aligned}\right\} \tag{4·2}$$

($i \neq j$, $i, j = 1, \ldots, t$; $u = 1, \ldots, v$). The conditions in (4·1) are imposed partially for convenience and partially to ensure determinacy in certain systems of equations which will be encountered. These numbers $\pi_{1u}, \ldots, \pi_{tu}$, will be considered as true ratings (or preferences) of the t objects $a_1, \ldots, a_t$ corresponding to the judge u.

5. THE LIKELIHOOD FUNCTION

From (4·2), the probability of judge u assigning rank r_{iju} to object a_i and rank $r_{jiu} = 3 - r_{iju}$ to object a_j upon comparison of the pair of objects is

$$n_{iju}\left(\frac{\pi_{iu}}{\pi_{iu}+\pi_{ju}}\right)^{2-r_{iju}}\left(\frac{\pi_{ju}}{\pi_{iu}+\pi_{ju}}\right)^{2-r_{jiu}} = n_{iju}\frac{\pi_{iu}^{2-r_{iju}}\pi_{ju}^{2-r_{jiu}}}{\pi_{iu}+\pi_{ju}}. \tag{5·1}$$

Provided that the pair of objects is compared by judge u, we observe that if $(a_i \rightarrow a_j \mid u)$, then $r_{iju} = 1$, $r_{jiu} = 2$ and (5·1) becomes

$$n_{iju}\frac{\pi_{iu}}{\pi_{iu}+\pi_{ju}} = P(a_i \rightarrow a_j \mid u),$$

and if $(a_i \leftarrow a_j \mid u)$, then $r_{iju} = 2$, $r_{jiu} = 1$ and (5·1) becomes

$$n_{iju}\frac{\pi_{ju}}{\pi_{iu}+\pi_{ju}} = P(a_i \leftarrow a_j \mid u).$$

If we assume probability independence between pairs of objects when compared by judge u, we obtain the likelihood function for the set of comparisons made by judge u,

$$L_u = \prod_R n_{iju} \frac{\pi_{iu}^{2-r_{iju}} \pi_{ju}^{2-r_{jiu}}}{\pi_{iu}+\pi_{ju}} = \frac{\prod_i \pi_{iu}^{2\alpha-\sum_j n_{iju} r_{iju}}}{\prod_R n_{iju}(\pi_{iu}+\pi_{ju})} \quad (u = 1, \ldots, v), \tag{5·2}$$

where R denotes the set of numbers $R = \{i, j: i < j, i = 1, \ldots, j-1; j = 2, \ldots, t; n_{iju} = 1\}$. If we assume probability independence between experiments performed by different judges, then from (5·2) we obtain the likelihood function for the complete set of comparisons made by the v judges,

$$L_{(1)} = \prod_u L_u, \tag{5·3}$$

and hence, if we set

$$a_i^* = 2\alpha v - \sum_j \sum_u n_{iju} r_{iju},$$

we have

$$\ln L_{(1)} = \sum_i a_i^* \ln \pi_{iu} - \sum_{i<j} \sum_u n_{iju} \ln (\pi_{iu}+\pi_{ju}). \tag{5·4}$$

6. Likelihood ratio tests

Test I. Consider the situation where each of the v judges compares only one set of r pairs, the set being specified by the paired comparison design used. We desire information about the differences, if any, among the true ratings of the t objects. Initially, we are willing to make the assumption that the v judges are consistent as a group; that is, we desire to test the null hypothesis,

$$H_0: \pi_{iu} = 1/t,$$

against the alternative hypothesis,

$$H_1: \pi_{iu} = \pi_i \quad (i = 1, \ldots, t; u = 1, \ldots, v), \tag{6·1}$$

where the π_i's are not all equal.

When the alternative hypothesis H_1 is true, (5·4) becomes

$$(\ln L_{(1)} \mid H_1) = \sum_i a_i^* \ln \pi_i - k \sum_{i<j} \ln (\pi_i + \pi_j), \tag{6·2}$$

where the simplification of the latter term is aided by the properties of the incidence matrices listed in § 3.

We note that a_i^*, defined in (5·4), can be rewritten, using (2·1), as

$$a_i^* = 2k(t-1) - \sum_j \sum_u n_{iju} r_{iju}.$$

When $r = \frac{1}{2}t(t-1)$, then $n_{iju} = 1$ for all i, j, u $(i \neq j)$, $k = v$, and

$$a_i^* = 2v(t-1) - \sum_j' \sum_u r_{iju},$$

and corresponds to a_i, defined by Bradley & Terry (1952), with n replaced by v.

Now maximizing (6·2), subject to the constraint $\sum_i \pi_i = 1$, yields the set of equations

$$\left.\begin{aligned} \frac{a_i^*}{\pi_i} - k \sum_j' (\pi_i + \pi_j)^{-1} + \mu = 0 \quad (i = 1, \ldots, t), \\ \sum_i \pi_i = 1. \end{aligned}\right\} \tag{6·3}$$

Summing (6·3) with respect to i yields

$$(2\alpha t - 4r)v + \mu = 0,$$

which implies $\mu = 0$, since from (2·1), $r = \frac{1}{2}t\alpha$. Hence the maximum-likelihood estimates $p_1, \ldots, p_t$ of $\pi_1, \ldots, \pi_t$ will be obtained from the system of equations

$$\left.\begin{aligned} \frac{a_i^*}{p_i} - k\sum_j{}' (p_i + p_j)^{-1} = 0 \quad (i = 1, \ldots, t), \\ \sum_i p_i = 1. \end{aligned}\right\} \tag{6·4}$$

When the null hypothesis H_0 is true, (5·4) becomes

$$\begin{aligned} (\ln L_{(1)} \mid H_0) &= (2\alpha vt - 3rv)\ln\frac{1}{t} - \frac{kt(t-1)}{2}\ln\frac{2}{t} \\ &= -\frac{kt(t-1)}{2}\ln 2 \\ &= -vr\ln 2. \end{aligned} \tag{6·5}$$

Thus the likelihood-ratio test of the null hypothesis H_0 against the alternative hypothesis H_1, as specified in (6·1), will be in terms of the likelihood ratio λ_1, where, if we define

$$B^{(1)} = k\sum_{i<j}\log(p_i + p_j) - \sum_i a_i^* \log p_i, \tag{6·6}$$

where $p_1, \ldots, p_t$ are solutions of (6·4), then

$$\ln\lambda_1 = -\{vr\ln 2 - B^{(1)}\ln 10\}. \tag{6·7}$$

When $r = \frac{1}{2}t(t-1)$, then $k = v$, and $B^{(1)}$ corresponds with the statistic B_1 discussed by Bradley & Terry (1952).

For the likelihood ratio in (6·7) to be useful to test the hypothesis stated in (6·1), it would be desirable to have some knowledge of the distribution of $B^{(1)}$—at least when the null hypothesis H_0 is true.

It is possible to generate all combinations of the object sums of ranks. Then under the null hypothesis of equality of true object ratings for each judge, the probability of each combination of rank sums is obtainable. This information would be sufficient to obtain the distribution of $B^{(1)}$ under H_0.

However, a direct computation of the probabilities of these various combinations of rank sums would be extremely tedious, even for small values of the design parameters. The fact that every permutation of the rank sums corresponding to each judge is not possible, since each judge does not compare all possible $\frac{1}{2}t(t-1)$ pairs, is the cause of the complication. However, we can utilize the symmetry of the designs to circumvent this particular difficulty.

Let A_i $(i = 1, \ldots, t)$ denote the set of $v\alpha = k(t-1)$ elements which are the $k(t-1)$ ranks assigned by all the v judges to the object a_i when compared with the other $(t-1)$ objects. That there are $v\alpha = k(t-1)$ such ranks follows from the fact that each of the v judges compares a_i with exactly α of the remaining $(t-1)$ objects. Also, from the design properties depicted in §§ 2 and 3, it follows that these $k(t-1)$ ranks are those for a_i after comparison with each of the remaining $(t-1)$ objects exactly k times. Hence, we can subdivide A_i into k disjoint subsets $A_{iu'}$, with each $A_{iu'}$ containing $(t-1)$ elements which are the ranks for

a_i after comparison with each of the remaining $(t-1)$ objects exactly once. Thus if we denote the $(t-1)$ elements of $A_{iu'}$ by $r_{iju'}$ $(i \neq j; j = 1, \ldots, t)$, then the sum of elements of $A_{iu'}$ is

$$S_{iu'} = \sum_j{}' r_{iju'}, \tag{6·8}$$

and the sum of the elements of A_i is

$$\sum_{u'=1}^{k} \sum_j{}' r_{iju'} = \sum_u \sum_j n_{iju} r_{iju}. \tag{6·9}$$

Hence the set of rank sums

$$\begin{aligned}(\sum_u \sum_j n_{1ju} r_{1ju}, \ldots, \sum_u \sum_j n_{tju} r_{tju}) &= \left(\sum_{u'=1}^{k} \sum_j{}' r_{1ju'}, \ldots, \sum_{u'=1}^{k} \sum_j{}' r_{tju'} \right) \\ &= \sum_{u'=1}^{k} (S_{1u'}, \ldots, S_{tu'}),\end{aligned} \tag{6·10}$$

could be considered as being the set of rank sums obtained from k complete repetitions of all $\frac{1}{2}t(t-1)$ pairs. It is these rank sums which determine $B^{(1)}$, and hence its distribution can be determined from the distribution of the different combinations of these rank sums. The possibility of putting the rank sums in the form of (6·10) enables the distribution of $B^{(1)}$, under the null hypothesis H_0 of equality of true object ratings, to be obtained in a manner described by Bradley & Terry (1952). Hence, tables presented by Bradley & Terry (1952) and by Bradley (1954*b*) may be used to provide the distribution of $B^{(1)}$ under the null hypothesis H_0. The tables are available for design parameters t and k in the following range: $t = 3, k = 1, \ldots, 10; t = 4, k = 1, \ldots, 8; t = 5, k = 1, \ldots, 5$. These tables also list the estimates, $p_1, \ldots, p_t$, corresponding to the rank sum combination (6·10).

Thus a test procedure for (6·1), for t and k in the indicated range, is as follows: Determine the rank sums (6·10). Then from the previously indicated tables, obtain the corresponding values for $p_1, \ldots, p_t$, $B^{(1)}$, and the probability P that $B^{(1)}$ will not be exceeded if the null hypothesis H_0 is true.

If either t or k is outside the range for the tables, the above test procedure cannot be used. However, if only k is outside the indicated range, it is possible to use the available tables to obtain the estimates $p_1, \ldots, p_t$, or at least a good first approximation to them, depending on whether or not there exists an integer c which divides the rank sums and k evenly and which is such that k/c is within the above indicated range for k. This technique is discussed in detail by Terry *et al.* (1952) and by Bradley (1954*b*).

If $t > 5$, the available tables will not be of assistance in obtaining approximations to $p_1, \ldots, p_t$ except in special cases. Hence, to obtain estimates, $p_1, \ldots, p_t$, in this case, it will be necessary to solve equations (6·4). Some methods aiding in the solution of equations of this form are suggested by Dykstra (1956), and Bradley & Terry (1952).

Once $p_1, \ldots, p_t$ have been obtained, $B^{(1)}$ can be evaluated from (6·6), but the significance level of $B^{(1)}$ can only be approximately determined. When this is the case, the test will be in terms of the statistic

$$T^{(1)} = -2 \ln \lambda_1 = 2vr \ln 2 - 2B^{(1)} \ln 10, \tag{6·11}$$

the discussion of which is reserved for § 7.

Test II. Suppose that in (6·1) the π_i's of the alternative hypothesis are divided into two groups, the elements within each group being equal. That is, we desire to test the null hypothesis

$$H_0\colon \pi_{iu} = 1/t \quad (i = 1, \ldots, t;\ u = 1, \ldots, v)$$

against the alternative hypothesis

$$H_2\colon \pi_{iu} = \left\{ \begin{array}{ll} \pi & (i = 1, \ldots, s), \\ \dfrac{1-s\pi}{t-s} & (i = s+1, \ldots, t),\ (u = 1, \ldots, v). \end{array} \right\} \tag{6·12}$$

For this situation, the equations (6·4) can be solved explicitly for p, the maximum-likelihood estimate of π, which, using (2·1) is

$$p = \frac{ks(4t-s-3) - 2\sum_{i=1}^{s}\sum_{j}\sum_{u} n_{iju}r_{iju}}{ks(5st-2t^2-6s+3t) - 2(2s-t)\sum_{i=1}^{s}\sum_{j}\sum_{u} n_{iju}r_{iju}}. \tag{6·13}$$

However, this expression diminishes in importance if we proceed in a manner analogous to that used by Bradley (1955) and let X denote the number of times an object of the first group of s objects ranks above an object of the second group of $(t-s)$ objects. Then

$$p = \frac{X}{ks(t-s)^2 + (2s-t)X}, \tag{6·14}$$

and the test procedure for (6·12) reduces to that of the binomial or sign test based on the $ks(t-s)$ comparisons of the objects of the first group with objects of the second group.

Test III. Suppose we have g groups with v judges in each group. We are interested in investigating the equality of the true treatment ratings of the t objects under the assumption of within-group judge consistency, but not necessarily assuming between-group judge consistency. It should be observed that these g groups could contain the same v judges, but due to some additional feature, such as significant time lag between repetitions, or training gained from continued experimentation, we are unwilling to assume between-group judge consistency.

Let $\pi_{1u}^m, \ldots, \pi_{tu}^m$ represent the true ratings of the t objects corresponding to the uth judge in the mth group $(u = 1, \ldots, v;\ m = 1, \ldots, g)$. Then, we wish to test the null hypothesis,

$$H_0\colon \pi_{iu}^m = 1/t,$$

against the alternative hypothesis,

$$H_3\colon \pi_{iu}^m = \pi_i^m \quad (i = 1, \ldots, t;\ u = 1, \ldots, v;\ m = 1, \ldots, g). \tag{6·15}$$

Let $n_{iju}^m = 1$ or 0 depending on whether or not the ith and jth objects are compared by the uth judge of the mth group $(i \neq j;\ i, j = 1, \ldots, t;\ u = 1, \ldots, v;\ m = 1, \ldots, g)$, and conventionally take $n_{iiu}^m = 0$ $(i = 1, \ldots, t;\ u = 1, \ldots, v;\ m = 1, \ldots, g)$. It is clear that the corresponding incidence matrices

$$N_u^m = (n_{iju}^m) \quad (m = 1, \ldots, g;\ u = 1, \ldots, v),$$

will have analogous properties to those given in (3·4).

Let r_{iju}^m be the rank assigned to the ith object if compared with the jth object by the uth judge of the mth group. Then, corresponding to (3·2), we have

$$r_{iju}^m + r_{jiu}^m = 3 \quad (m = 1, \ldots, g;\ u = 1, \ldots, v),$$

when the ith and jth objects are compared by the uth judge of the mth group.

Now, under the assumption of probability independence between pairs of objects, judges and groups of judges, and by defining

$$a^*_{im} = 2\alpha v - \sum_j \sum_u n^m_{iju} r^m_{iju}, \tag{6·16}$$

we obtain

$$(\ln L_{(3)} \mid H_3) = \sum_m \{\sum_i a^*_{im} \ln \pi^m_i - k \sum_{i<j} \ln (\pi^m_i + \pi^m_j)\}, \tag{6·17}$$

$$(\ln L_{(3)} \mid H_0) = -vrg \ln 2, \tag{6·18}$$

where $L_{(3)}$ is the likelihood function, and $(\ln L_{(3)} \mid H_i)$ denotes the natural logarithm of the likelihood function when the hypothesis H_i is true ($i = 0, 3$). Then, if we denote the likelihood ratio by λ_3 and define

$$B^{(1)}_m = k \sum_{i<j} \log (p^m_i + p^m_j) - \sum_i a^*_{im} \log p^m_i, \tag{6·19}$$

where $p^m_1, \ldots, p^m_t$, the maximum-likelihood estimates of $\pi^m_1, \ldots, \pi^m_t$, are solutions of the system of equations

$$\left.\begin{aligned} \frac{a^*_{im}}{p^m_i} - k \sum_j{}' (p^m_i + p^m_j)^{-1} = 0 \quad (i = 1, \ldots, t), \\ \sum_i p^m_i = 1 \quad (m = 1, \ldots, g), \end{aligned}\right\} \tag{6·20}$$

and define

$$B^{(3)} = \sum_m B^{(1)}_m, \tag{6·21}$$

and

$$T^{(1)}_m = 2vr \ln 2 - 2B^{(1)}_m \ln 10, \tag{6·22}$$

we obtain

$$T^{(3)} = -2 \ln \lambda_3 = \sum_m T^{(1)}_m. \tag{6·23}$$

The probability of a specified value B of $B^{(3)}$ can be obtained by determining the probabilities of the joint occurrences of all combinations of $B^{(1)}_1, \ldots, B^{(1)}_g$ such that $\sum_m B^{(1)}_m = B$, and then summing these probabilities. Clearly, the distribution of $B^{(1)}_m$ ($m = 1, \ldots, g$) under H_0 is the same as the distribution of $B^{(1)}$ under H_0. Hence, the distribution of $B^{(3)}$ under H_0 can be obtained from the distribution of $B^{(1)}$ under H_0. Tables of this distribution for $t = 3, g$, $k = 2, \ldots, 5$ are given by Bradley & Terry (1952), and for $t = 4, g, k = 2$ by Wilkinson (1956).

The test procedure for (6·15) is then the following: Compute the rank sums

$$\sum_j \sum_u n^m_{iju} r^m_{iju} \quad (i = 1, \ldots, t),$$

corresponding to the mth group of judges ($m = 1, \ldots, g$). Then, if the design parameters are within the range $t = 3$, $k = 1, \ldots, 10$; $t = 4$, $k = 1, \ldots, 8$; $t = 5$, $k = 1, \ldots, 5$, determine $B^{(1)}_m$ from the previously indicated tables. Sum these $B^{(1)}_m$'s with respect to m, obtaining $B^{(3)}$. If $t = 3, g, k = 2, \ldots, 5$; $t = 4, g, k = 2$, the exact significance level is given in the above-mentioned tables. Otherwise, without further extension of these tables, an exact test of (6·15) will not be available. When this is the case, the test will be in terms of the statistic $T^{(3)}$ of (6·23), the approximate significance level of which can be determined as indicated in § 7.

For large t, g, k, it will be necessary to solve the equations in (6·20) for $p^m_1, \ldots, p^m_t$ ($m = 1, \ldots, g$), and then to evaluate $B^{(3)}$ using (6·19) and (6·21).

Test IV. It is conceivable that the g groups of v judges considered in Test III could have had the same judges in each group. For this situation, if we are willing to assume consistency of the judges from one repetition to the next, we may wish to test the null hypothesis,

$$H_0\colon \pi_{iu}^m = 1/t,$$

against the alternative hypothesis,

$$H_4\colon \pi_{iu}^m = \pi_i \quad (i = 1, \ldots, t;\ u = 1, \ldots, v;\ m = 1, \ldots, g). \tag{6·24}$$

The likelihood ratio test for (6·24) is provided by computing the rank sums

$$\sum_m \sum_j \sum_u n_{ijm}^m r_{iju}^m \quad (i = 1, \ldots, t),$$

and

$$B^{(4)} = gk \sum_{i<j} \log (p_i + p_j) - \sum_i a_i^1 \log p_i, \tag{6·25}$$

where $a_i^1 = \sum_m a_{im}^*$ and $p_1, \ldots, p_t$, the maximum-likelihood estimates of $\pi_1, \ldots, \pi_t$, are solutions of the equations

$$\frac{a_i^1}{p_i} - gk \sum_j{}' (p_i + p_j)^{-1} = 0 \quad (i = 1, \ldots, t),$$

$$\sum_i p_i = 1.$$

$B^{(4)}$ corresponds to $B^{(1)}$ with k replaced by gk. Thus, for t and gk within the range of available tables, $B^{(4)}$ and its exact significance level can be obtained directly from these tables. When either t or gk is outside the range of available tables, an exact test is not available. In this case, the test will be in terms of the statistic

$$T^{(4)} = 2vrg \ln 2 - 2B^{(4)} \ln 10, \tag{6·26}$$

the approximate significance level of which can be determined as indicated in § 7.

It should be noted that there are many ways in which g sets of r pairs could be assigned to each of the judges. No matter how the sets of pairs are assigned, the statistic $B^{(4)}$ will have the same distribution under the null hypothesis H_0 of (6·24). This would suggest that simplicity of assignment of the g sets to the v judges would be the popular criterion. However, it is felt that it would be better to have each judge compare as many different sets of r pairs as possible, as this may provide greater connectivity. The investigation of an exact criterion is desirable, but will not be considered at this time.

Test V. For the situation of Test III we now wish to test for judge consistency between groups; that is, we wish to test the null hypothesis,

$$H_4\colon \pi_{iu}^m = \pi_i,$$

against the alternative hypothesis,

$$H_3\colon \pi_{iu}^m = \pi_i^m \quad (i = 1, \ldots, t;\ u = 1, \ldots, v;\ m = 1, \ldots, g).$$

The likelihood-ratio test is in terms of the statistic

$$T^{(5)} = 2(B^{(4)} - B^{(3)}) \ln 10, \tag{6·27}$$

the distribution of which under the null hypothesis H_4 will depend on $\pi_1, \ldots, \pi_t$. Hence, $T^{(5)}$ will not provide an exact parameter-free test, but will supply an approximate test as discussed in § 7.

It is natural to desire a test concerning the true ratings of the t objects without assuming consistency of the judges as a group; that is, we would like to have a test of the null hypothesis,

$$H_0\colon \pi_{iu} = 1/t,$$

against the alternative hypothesis,

$$H\colon \text{the } \pi_{iu}\text{'s are not all equal.}$$

The likelihood ratio could be determined but its distribution would depend on the paired comparison design being used, the number g of repetitions made by each judge, and the criterion for assigning the g sets of r pairs to each judge. The assignment of these sets is, in itself, a design problem. However, it is conceivable that with the assignment of these sets to satisfy certain additional symmetry conditions, tabulation of the distribution for small values of t would be practicable.

Of course, if $g = cv$, $c \geqslant 1$, such that each judge compares each of all the possible v different sets of r pairs c times, the situation reduces to the equivalent of each judge making the comparison of ck sets of all possible $\frac{1}{2}t(t-1)$ pairs. A test for this is given by Bradley & Terry (1952).

7. Large-sample distributions

Let
$$y_i = t\left(p_i - \frac{1}{t}\right) \quad (i = 1, \ldots, t), \tag{7·1}$$

where $p_1, \ldots, p_t$ are the maximum-likelihood estimates of $\pi_1, \ldots, \pi_t$ given in (6·1). Now observe that

$$\sum_i \sum_j \sum_u n_{iju} r_{iju} = \tfrac{3}{2}kt(t-1);$$

hence,
$$\begin{aligned} \sum_i a_i^* &= 2\alpha vt - \tfrac{3}{2}kt(t-1) \\ &= \tfrac{1}{2}kt(t-1), \end{aligned}$$

where a_i^* is defined in (5·4). Now, upon substitution in (6·6) and (6·11), we obtain

$$B^{(1)} = \tfrac{1}{2}kt(t-1)\log 2 + k\sum_{i<j} \log\{1 + \tfrac{1}{2}(y_i + y_j)\} - \sum_i a_i^* \log(1 + y_i), \tag{7·2}$$

and
$$T^{(1)} = 2k\sum_{i<j} \ln\{1 + \tfrac{1}{2}(y_i + y_j)\}^{-1} + 2\sum_i a_i^* \ln(1 + y_i), \tag{7·3}$$

respectively. Substitution of (7·1) in (6·4) yields

$$a_i^* = \tfrac{1}{2}k(1 + y_i)\sum_j{}' \{1 + \tfrac{1}{2}(y_i + y_j)\}^{-1},$$

and hence, upon substitution for a_i^* in (7·3), we obtain

$$T^{(1)} = k\sum_i (1 + y_i)\ln(1 + y_i)\sum_j{}' \{1 + \tfrac{1}{2}(y_i + y_j)\}^{-1} + 2k\sum_{i<j} \ln\{1 + \tfrac{1}{2}(y_i + y_j)\}^{-1}. \tag{7·4}$$

$T^{(1)}$, as expressed in (7·4), can be put in the form

$$T^{(1)} = \tfrac{1}{4}kt\sum_i y_i^2 + R(y_i), \tag{7·5}$$

where $R(y_i)$ depends on higher powers of y_i than the second power. Then, in the manner followed by Bradley (1955), if we redefine

$$\pi_i = \frac{1}{t} + \frac{\delta_{ik}}{\sqrt{k}} \quad (i = 1, \ldots, t),$$

where δ_{ik} is a sequence of constants converging to δ_i as $k \to \infty$, it can be shown that $R(y_i)$ converges to zero in probability as $k \to \infty$, and that $T^{(1)}$ has the same limiting distribution as $\frac{1}{4}kt \sum_i y_i^2$. This limiting distribution is, under H_1, a non-central χ^2-distribution with $(t-1)$ degrees of freedom and parameter of non-centrality

$$\lambda = \tfrac{1}{4}t^3 \sum_i \delta_i^2,$$

which, for large k, can be approximated by

$$\lambda = \tfrac{1}{4}kt^3 \sum_i \left(\pi_i - \frac{1}{t}\right)^2.$$

Under the null hypothesis H_0, $\pi_i = 1/t$, $\delta_{ik} = \delta_i = 0$ $(i = 1, \ldots, t)$. Therefore $\lambda = 0$ and thus, under H_0, $T^{(1)}$ has a limiting central χ^2-distribution with $(t-1)$ degrees of freedom.

Similarly, it can be shown that $T^{(3)}$, $T^{(4)}$ and $T^{(5)}$ have, under their respective null hypotheses, limiting central χ^2-distributions with $g(t-1)$, $(t-1)$ and $(g-1)(t-1)$ degrees of freedom, respectively. Thus approximate significance levels for the indicated tests can be obtained from tables for the χ^2-distribution with appropriate numbers of degrees of freedom.

8. EXTREME SETS OF RANK SUMS

For the situation in Test I, the extreme values which the sum of ranks corresponding to the ith object may have are

$$\sum_j \sum_u n_{iju} r_{iju} = 2\alpha v \quad \text{or} \quad \alpha v \quad (i = 1, \ldots, t).$$

When either of these cases arises for any of the objects, equations in (6·4), which provide our maximum-likelihood estimates for the parameters, become somewhat valueless. In order for this rank sum to be $2\alpha v$ (or αv), it is necessary that the ith object be judged inferior (or superior) in all comparisons made by all judges. When these situations arise, the procedure is to estimate π_i by 0 (or 1); then, drop this object from the analysis and consider only the remaining $t-1$ objects.

Let us consider the case where the tth object has the extreme sum of ranks

$$\sum_j \sum_u n_{tju} r_{tju} = 2\alpha v.$$

Omitting this object from consideration, we will have a reduced set of $(t-1)$ rank sums, namely,

$$\sum_j \sum_u n_{1ju} r_{1ju} - k, \quad \ldots, \quad \sum_j \sum_u n_{t-1,ju} r_{t-1,ju} - k. \tag{8·1}$$

Now, from the symmetry of the paired comparison designs, we obtain the logarithm of the likelihood function $L_{(1)}$ under the alternative hypothesis H_1 of (6·1) to be

$$(\ln L_{(1)} \mid H_1) = \sum_{i=1}^{t-1} \left\{ 2k(t-2) - \sum_{j=1}^{t-1} \sum_u n_{iju} r_{iju} \right\} \ln \pi_i - k \sum_{i<j}^{t-1} \ln (\pi_i + \pi_j).$$

It also follows easily that the system of equations, which provides maximum-likelihood estimates of $\pi_1, \ldots, \pi_{t-1}$, and the test statistic $B^{(1)}$ will be the same as in (6·4) and (6·6), respectively, with t replaced by $t-1$.

Hence, the test procedure for (6·1) will be the same as that previously described with t replaced by $t-1$, and with the original set of rank sums replaced by the reduced set (8·1).

For the case where the tth object has the extreme rank sum

$$\sum_j \sum_u n_{tju} r_{tju} = \alpha v,$$

the omission of this object from consideration will result in the reduced set of $(t-1)$ rank sums

$$\sum_j \sum_u n_{1ju} r_{1ju} - 2k, \quad \ldots, \quad \sum_j \sum_u n_{t-1,ju} r_{t-1,ju} - 2k. \tag{8·2}$$

Similarly, the test procedure for (6·1), in this case, is the same as that previously described, with t replaced by $(t-1)$, and with the original set of rank sums replaced by the reduced set (8·2).

In Test IV, for the case where the tth object has the extreme rank sum

$$\sum_m \sum_j \sum_u n^m_{tju} r^m_{tju} = 2\alpha g v \quad \text{or} \quad \alpha g v, \tag{8·3}$$

the test procedure for (6·24) is the same as previously described, with t replaced by $(t-1)$, and with the original set of rank sums replaced by

$$\left.\begin{aligned} &\sum_m \sum_j \sum_u n^m_{1ju} r^m_{1ju} - gk, \quad \ldots, \quad \sum_m \sum_j \sum_u n^m_{t-1,ju} r^m_{t-1,ju} - gk, \\ \text{or} \quad &\sum_m \sum_j \sum_u n^m_{1ju} r^m_{1ju} - 2gk, \quad \ldots, \quad \sum_m \sum_j \sum_u n^m_{t-1,ju} r^m_{t-1,ju} - 2gk, \end{aligned}\right\} \tag{8·4}$$

depending on whether or not the tth object rank sum is $2\alpha g v$ or $\alpha g v$, respectively.

For Tests III and V, if an extreme rank sum occurs for the tth object of the mth group, namely,

$$\sum_j \sum_u n^m_{tju} r^m_{tju} = 2\alpha v \quad \text{or} \quad \alpha v,$$

the respective test procedures can be employed with t replaced by $(t-1)$, and with the original set of rank sums replaced by the reduced set of rank sums,

$$\begin{aligned} &\sum_j \sum_u n^m_{1ju} r^m_{1ju} - k, \quad \ldots, \quad \sum_j \sum_u n^m_{t-1,ju} r^m_{t-1,ju} - k, \\ \text{or} \quad &\sum_j \sum_u n^m_{1ju} r^m_{1ju} - 2k, \quad \ldots, \quad \sum_j \sum_u n^m_{t-1,ju} r^m_{t-1,ju} - 2k. \end{aligned}$$

Such an alteration to the test procedure can be extended when more than one of the g groups has an extreme rank sum. For Test V, if the tth object in all g groups has an extreme rank sum as in (8·3), in order to evaluate $B^{(4)}$ we will need to follow the procedure described in Test IV, with t replaced by $t-1$, and to employ the reduced set of rank sums of (8·4).

9. Examples to illustrate the procedures of analysis†

Example 1. Consider $t = 5$ specimens of handwriting compared pairwise by $v = 6$ judges, with each judge recording a preference for one object of each pair which he compares. We will refer to the factors which influence judgement as characteristic x. The r pairs compared

† The data used in these examples is a small portion of that collected for an experiment conducted at the Psychometric Laboratory of the University of North Carolina. It is used to illustrate the procedures of analysis only.

by the uth judge ($u = 1, \ldots, 6$) will be dictated by the field plan of design (2), Table 1, given by Bose (1956). The incidence matrices for this design are:

$$\left.\begin{aligned} N_1 &= \begin{pmatrix} 0 & 0 & 1 & 1 & 0 \\ 0 & 0 & 0 & 1 & 1 \\ 1 & 0 & 0 & 0 & 1 \\ 1 & 1 & 0 & 0 & 0 \\ 0 & 1 & 1 & 0 & 0 \end{pmatrix}, \quad N_2 = \begin{pmatrix} 0 & 0 & 0 & 1 & 1 \\ 0 & 0 & 1 & 0 & 1 \\ 0 & 1 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 & 0 \end{pmatrix}, \\ N_3 &= \begin{pmatrix} 0 & 1 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 1 & 0 \end{pmatrix}, \quad N_4 = \begin{pmatrix} 0 & 1 & 0 & 0 & 1 \\ 1 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 & 1 \\ 0 & 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 \end{pmatrix}, \\ N_5 &= \begin{pmatrix} 0 & 1 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 & 1 \\ 0 & 1 & 0 & 1 & 0 \end{pmatrix}, \quad N_6 = \begin{pmatrix} 0 & 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 1 & 0 \\ 1 & 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 1 \\ 1 & 0 & 0 & 1 & 0 \end{pmatrix}. \end{aligned}\right\} \tag{9·1}$$

Now, for judge u ($u = 1, \ldots, v$), we construct a $(t \times t)$ preference matrix $R_u = (n_{iju} r_{iju})$, where $r_{iju} = 1$ or 2 if $(a_i \rightarrow a_j \mid u)$ or $(a_i \leftarrow a_j \mid u)$, respectively. The preference matrix R_u will have 1's and 2's as elements corresponding to the unit elements of N_u, and zeros corresponding to the zero elements of N_u. From this definition of R_u, we observe that the sums of the rows of R_u yield a $(t \times 1)$ vector, the elements of which are the rank sums of the t objects corresponding to judge u. For example, the preference matrix R_1 and the rank sums vector corresponding to judge J_1 for this experiment are

$$R_1 = \begin{pmatrix} 0 & 0 & 2 & 1 & 0 \\ 0 & 0 & 0 & 1 & 1 \\ 1 & 0 & 0 & 0 & 2 \\ 2 & 2 & 0 & 0 & 0 \\ 0 & 2 & 1 & 0 & 0 \end{pmatrix},$$

$$R_1 l = \begin{pmatrix} 3 \\ 2 \\ 3 \\ 4 \\ 3 \end{pmatrix} = \begin{pmatrix} \sum_j n_{1j1} r_{1j1} \\ \sum_j n_{2j1} r_{2j1} \\ \sum_j n_{3j1} r_{3j1} \\ \sum_j n_{4j1} r_{4j1} \\ \sum_j n_{5j1} r_{5j1} \end{pmatrix},$$

where l is a unit column vector of appropriate dimension. By this procedure we obtain for each judge the following rank sums vectors:

$$\left.\begin{aligned} J_1&: \ (3, 2, 3, 4, 3), \\ J_2&: \ (2, 2, 4, 3, 4), \\ J_3&: \ (2, 4, 3, 4, 2), \\ J_4&: \ (2, 3, 3, 3, 4), \\ J_5&: \ (4, 2, 2, 4, 3), \\ J_6&: \ (2, 2, 4, 4, 3). \end{aligned}\right\} \tag{9·2}$$

Summing these with respect to judges yields

$$(\sum_j \sum_u n_{1ju} r_{1ju}, \ldots, \sum_j \sum_u n_{5ju} r_{5ju}) = (15, 15, 19, 22, 19). \tag{9·3}$$

Hence

$$(a_1^*, \ldots, a_5^*) = (9, 9, 5, 2, 5), \tag{9·4}$$

where a_i^* is defined in (5·4). Entering the tables of Bradley (1954*b*) for $t = 5$ and for $k = 3$ complete repetitions, we obtain, corresponding to the rank sums (9·3),

$$(p_1, \ldots, p_5) = (0{\cdot}38, 0{\cdot}38, 0{\cdot}10, 0{\cdot}03, 0{\cdot}10), \quad B^{(1)} = 6{\cdot}686,$$

$$P\{B^{(1)} \leqslant 6{\cdot}686 \mid H_0\} = 0{\cdot}0404.$$

Hence, we would conclude, at the 0·0404 level of significance, that the five handwriting specimens are different with respect to the characteristic x under the assumption that the judges are consistent as a group.

It is of interest to note, from (6·11), that

$$T^{(1)} = 2vr \ln 2 - 2B^{(1)} \ln 10 = 10{\cdot}80 \simeq \chi^2_{(4)}.$$

From the large-sample properties of $T^{(1)}$, we obtain the approximate significance level to be $P(\chi^2_{(4)} \geqslant 10{\cdot}80 \mid H_0) = 0{\cdot}028$. Thus, although the approximate test will give too many significant results, the approximate significance level obtained is reasonably close to the exact significance level, even for k as small as 3.

Example 2. Suppose that the six judges of Example 1 make two incomplete repetitions with r pairs in each repetition. The r pairs compared by the uth judge in the first repetition are those indicated in (9·1), and, hence, the incidence matrices for the first repetition are

$$N_u^1 = N_u \quad (u = 1, \ldots, 6).$$

The r pairs compared by the uth judge in the second repetition are indicated by the incidence matrices

$$N_u^2 = N_{u+1} \quad (u = 1, \ldots, 5), \quad N_6^2 = N_1.$$

The rank sums vectors corresponding to each judge for the first repetition are given in (9·2). By an analogous procedure to that used in Example 1, we compute the rank sums vectors corresponding to each judge on the second repetition:

$$\left.\begin{array}{ll} J_1\colon & (2, 3, 2, 4, 4), \\ J_2\colon & (3, 2, 3, 3, 4), \\ J_3\colon & (2, 3, 3, 4, 3), \\ J_4\colon & (2, 3, 3, 3, 4), \\ J_5\colon & (3, 2, 3, 4, 3), \\ J_6\colon & (2, 2, 4, 4, 3). \end{array}\right\} \tag{9·5}$$

Summing these with respect to judges, we obtain

$$(\sum_j \sum_u n_{1ju}^2 r_{1ju}^2, \ldots, \sum_j \sum_u n_{5ju}^2 r_{5ju}^2) = (14, 15, 18, 22, 21). \tag{9·6}$$

Hence

$$(a_{12}^*, \ldots, a_{52}^*) = (10, 9, 6, 2, 3), \tag{9·7}$$

where a_{12}^* is defined in (6·19). By summing the vectors in (9·6) and (9·3), we obtain

$$(\sum_m \sum_j \sum_u n_{1ju}^m r_{1jw}^m, \ldots, \sum_m \sum_j \sum_u n_{5ju}^m r_{5ju}^m) = (29, 30, 37, 44, 40). \tag{9·8}$$

Hence, we obtain $$(a_1^1, \ldots, a_5^1) = (19, 18, 11, 4, 8). \tag{9·9}$$

The result in (9·9) can be checked by adding vectors in (9·7) and (9·4). The elements of the vector (9·8) are the rank sums which enter into the statistic $B^{(4)}$ of (6·25). Here $t = 5$ and $gk = 6$—the corresponding number of complete repetitions—are outside the range of available tables. Hence we are unable to obtain either the value of $B^{(4)}$ or its significance level directly from the tables. However, since $t = 5$, we can use a procedure outlined by Bradley (1954*b*) and Terry *et al.* (1952) to assist in evaluating $B^{(4)}$. Dividing gk and the elements of (9·8) by two, we obtain

$$\tfrac{1}{2}gk = 3, \quad \tfrac{1}{2}(9·8) = (14·5, 15, 18·5, 22, 20). \tag{9·10}$$

Then, for rank sums vectors

$$(14, 15, 19, 22, 20) \quad \text{and} \quad (15, 15, 18, 22, 20),$$

we enter the tables for $t = 5$ and for three complete repetitions and obtain corresponding estimates

$$(0·51, 0·33, 0·08, 0·02, 0·05) \quad \text{and} \quad (0·38, 0·38, 0·14, 0·03, 0·07),$$

respectively. Then, by linear interpolation, we obtain an approximation to the estimates, $p_1, \ldots, p_5$, corresponding to the elements of (9·10) to be

$$(0·445, 0·355, 0·110, 0·025, 0·060),$$

which, after adjustment to add to 1, is

$$(0·447, 0·357, 0·111, 0·025, 0·060).$$

Starting with these approximations and using the iterative formula given by Bradley & Terry (1952), we obtain

$$(p_1, \ldots, p_5) = (0·441, 0·361, 0·106, 0·029, 0·063).$$

Substituting these values in (6·25), we obtain $B^{(4)} = 12·556$. This value substituted in (6·26) yields $T^{(4)} = 25·353$, and since $T^{(4)}$ has approximately a χ^2-distribution, under H_0, with $t-1 = 4$ degrees of freedom, it follows that $P(T^{(4)} \geqslant 25·353 \mid H_0) < 0·0005$, approximately. Hence, under the assumption that the judges are consistent as a group and consistent from one repetition to the next, we would conclude, at less than the 0·0005 level of significance (approximately), that the five handwriting specimens are different with respect to the characteristic x.

Example 3. Suppose that in the previous example we were unwilling to assume judge consistency from one repetition to the next. The situation then becomes that for Test III. Thus we have $g = 2$ groups of $v = 6$ judges comparing t objects according to a characteristic x, and, with reference to this characteristic, the judges in each group would not be considered the same.

Using the data of Examples 1 and 2 for each repetition separately, we obtain

$$(p_1, \ldots, p_5) = (0·38, 0·38, 0·10, 0·03, 0·10), \quad B_1^{(1)} = 6·686, \quad T_1^{(1)} = 10·800,$$

$$(p_1, \ldots, p_5) = (0·51, 0·33, 0·10, 0·02, 0·03), \quad B_2^{(1)} = 5·598, \quad T_2^{(1)} = 15·809.$$

Hence, from (6·23), $T^{(3)} = 10{\cdot}800 + 15{\cdot}809 = 26{\cdot}609$. Since $T^{(3)}$ has, under H_0, approximately a χ^2-distribution with $g(t-1) = 8$ degrees of freedom, an approximate significance level for Test III is $P(T^{(3)} \geqslant 26{\cdot}609 \mid H_0) = 0{\cdot}001$. Hence, under the assumption of within-group judge consistency only, we conclude, at approximately the 0·001 level of significance, that the five handwriting specimens are different with respect to the characteristic x.

Example 4. Suppose for the experiment used in Examples 2 and 3 we now desire to test for the consistency from one group of repetitions to the next. Then Test V is the one needed.

From (6·27) we have the test statistic

$$T^{(5)} = 2(B^{(4)} - B^{(3)}) \ln 10 = T^{(3)} - T^{(4)} = 1{\cdot}256.$$

Since $T^{(5)}$ has, under H_4, approximately a χ^2-distribution with $(g-1)(t-1) = 4$ degrees o freedom, the approximate significance level for Test V, in this case, is

$$P(T^{(5)} \geqslant 1{\cdot}256 \mid H_4) = 0{\cdot}86.$$

Hence, under the assumption of within-group judge consistency, we have no evidence to doubt the hypothesis that the group judging criterion is the same from one repetition to the next.

Example 5. Suppose that in Example 1 we are only interested in testing the null hypothesis—that the true object ratings are equal—against the alternative that the true ratings are in two groups, within which they are equal, but between which they are not necessarily equal. A test for this situation is given by Test II. For this example we will take $s = 2$ objects in the first group and $t - s = 3$ objects in the second group. To determine X, the number of times an object of the first group of $s = 2$ objects is ranked above an object of the second group of $t - s = 3$ objects, recode the objects in such a way that $a_1, \ldots, a_s$ are in the first group and $a_{s+1}, \ldots, a_t$ are in the second group. Then for judge u ($u = 1, \ldots, v$), construct a new preference matrix corresponding to R_u of Example 1, but this time use the notation

$$r_{iju} = \begin{cases} 1 & \text{if } (a_i \rightarrow a_j \mid u) \\ 0 & \text{if } (a_i \leftarrow a_j \mid u), \end{cases}$$

if objects a_i and a_j are compared by judge u. Then X will be the sum of the elements of the v, $(s \times \overline{t-s})$ submatrices contained in the upper right-hand corner of the new $t \times t$ preference matrices corresponding to each judge. For the data of our example, the six, (2×3) submatrices are:

$$\begin{pmatrix} 0 & 1 & 0 \\ 0 & 1 & 1 \end{pmatrix}, \quad \begin{pmatrix} 0 & 1 & 1 \\ 1 & 0 & 1 \end{pmatrix}, \quad \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 0 \end{pmatrix},$$

$$\begin{pmatrix} 0 & 0 & 1 \\ 0 & 1 & 0 \end{pmatrix}, \quad \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix}, \quad \begin{pmatrix} 1 & 0 & 1 \\ 1 & 1 & 0 \end{pmatrix}.$$

Summing the elements of these, we obtain $X = 15$. Hence, from (6·14), the maximum-likelihood estimate of π is

$$p = \frac{X}{ks(t-s)^2 + (2s-t)\,X} = 0{\cdot}385.$$

If we consider the preference of an object in the first group over an object in the second group as a success, we have $X = 15$ successes in $ks(t-s) = 18$ trials. The probability of such an occurrence, under the hypothesis that the probability is one-half that an object of the

first group is ranked over an object of the second group, is $P(X = 15 \mid H_0) = 0{\cdot}0031$. In addition, $P(X \geqslant 15 \mid H_0) = 0{\cdot}0096$. Hence, under the assumption that the judges are consistent as a group, we would reject the null hypothesis of equality of true object ratings in favour of the alternative hypothesis H_2, at approximately the 0·01 level of significance.

The author wishes to thank Prof. R. C. Bose for his suggestion of the problem and his helpful guidance throughout its investigation.

REFERENCES

BOSE, R. C. (1956). Paired comparison designs for testing concordance between judges. *Biometrika*, **43**, 113.

BRADLEY, R. A. & TERRY, M. E. (1952). Rank analysis of incomplete block designs. I. The method of paired comparisons. *Biometrika*, **39**, 324.

BRADLEY, R. A. (1954*a*). Incomplete block rank analysis: On appropriateness of the model for a method of paired comparisons. *Biometrics*, **10**, 375.

BRADLEY, R. A. (1954*b*). Rank analysis of incomplete block designs. II. Additional tables for the method of paired comparisons. *Biometrika*, **41**, 502.

BRADLEY, R. A. (1955). Rank analysis of incomplete block designs. III. Some large-sample results on estimation and power for a method of paired comparisons. *Biometrika*, **42**, 450.

DYKSTRA, OTTO, Jr. (1956). A note on the rank analysis of incomplete block designs; applications beyond the scope of existing tables. *Biometrics*, **12**, 301.

KENDALL, M. G. (1955). Further contributions to the theory of paired comparisons. *Biometrics*, **11**, 43.

TERRY, M. E., BRADLEY, R. A. & DAVIS, L. L. (1952). New designs and techniques for organoleptic testing. *Food Tech.* **6**, 250.

WILKINSON, J. W. (1956). Analysis of paired comparison designs with incomplete repetitions. Institute of Statistics, University of North Carolina, Mimeo. Series 147.

NON-NULL RANKING MODELS. I

By C. L. MALLOWS
University College, London

1. Introduction and summary

Kendall (1950) has remarked that the major outstanding problem in ranking theory is the specification of a suitable population of ranks in non-null cases. Much attention has been concentrated on situations which Daniels (1950, § 5) calls of type (i):

'The sample is regarded as having been randomly chosen from a bivariate population of ranks'

the underlying population being either finite or infinite, e.g. bivariate Normal. Rather less work has been done on Daniels's type (ii):

'There is a fixed set of individuals being assessed by a population of judges, or by the same judge in repeated trials, on a particular attribute whose ranking is known *a priori*. The random element is uncertainty of preference, the correlation being the result of real differences between the individuals, and the population is one of rankings conditional on a given objective order.'

Daniels (1950), following Babington Smith, Thurstone and Mann, treats this problem as one of regression. The present approach is by way of paired-comparison theory. The judge is assumed to arrive at a ranking of n objects $U_1, U_2, \ldots, U_n$ by first making all the ${}^nC_2$ comparisons between pairs *independently*, but then only accepting the results if they are consistent with a ranking of the n objects.

Various non-null models are proposed. The general model (eqn. (1)) depends on ${}^nC_2$ parameters—this number is then reduced to $n-1$ by using the Bradley–Terry (1952) paired-comparison model (eqn. (4)). An alternative method of simplification is proposed which makes the probability of putting (in paired-comparisons) any two objects U_i and U_j in the correct order equal to

$$\tfrac{1}{2}+\tfrac{1}{2}\tanh(k\log\theta+\log\phi),$$

where θ and ϕ are parameters, and $k=j-i$ is the difference between the true ranks of the two objects. Thus this probability is a simple monotonic function of a quantity which is composed of a term increasing linearly with k, and a constant term. The null hypothesis corresponds to $\theta=\phi=1$. It is found that θ is associated with Spearman's coefficient r_s, and ϕ with Kendall's t_k (eqn. (9)). Each of these parameters has a further interpretation. Thus θ may be regarded as assigning weights to the n objects, these weights being in geometric progression. The paired-comparisons are then made in such a way that the probability of ranking one object higher than another is a simple function of the weights assigned to these two objects. The parameter ϕ has the following further interpretation: having obtained a ranking of length $n-1$ and wishing to introduce a further object, ϕ specifies the probabilities of this object being ranked in the various possible positions; these probabilities being in geometric progression, decreasing away from the true position (eqn. (18)).

Putting $\phi=1$ in the general model gives a special case of the Bradley–Terry model (eqn. (11)). It is shown, however, that asymptotically, when the joint distribution of r_s

and t_k tends to the bivariate Normal form, the two coefficients cannot distinguish between the two parameters; it is therefore proposed to put $\theta = 1$ and to use only the one parameter ϕ (eqn. (12)). This leads to exceptionally simple results, including an explicit form for the probability generating function (p.g.f.) of t_k and an invariance property of the probabilities p_{ij} of obtaining two objects U_i and U_j in the correct order in the ranking; p_{ij} is found to depend only on $j-i$ and ϕ, and not on i or n (§ 9).

Methods of estimating ϕ are given; tests are derived which may be used to decide whether two judges' rankings of the same objects are consistent with the same non-null hypothesis; more generally, given m judges, each of whom produces l rankings, we may test both for differences between judges and for inconsistencies within judges. The power of these tests follows immediately from standard theory.

An expression is derived giving the conditional expectation of r_s for given t_k.

2. THE GENERAL MODEL, $\mathscr{H}_\lambda$

In the method of paired-comparisons, n objects $U_1, U_2, \ldots, U_n$, whose *a priori* ranking is here assumed known, are ranked in pairs, there being ${}^nC_2$ such comparisons in all. Babington Smith (1950) suggested that a suitable non-null hypothesis for this situation would be to assign to each pair (i, j) $(i<j)$ the probability π_{ij} of ranking U_i lower than U_j; this we shall denote by

$$\pi_{ij} = P\{U_i \prec U_j\} \quad (1 \leqslant i < j \leqslant n),$$

and to assume that all the comparisons are independent.

The proposed general non-null ranking model, depending on ${}^nC_2$ parameters π_{ij}, is as follows:

(i) The above model is used to generate a set of ${}^nC_2$ comparisons.

(ii) If the resulting complex of comparisons is consistent, i.e. if there are no circular triads of comparisons such as $U_i \prec U_j \prec U_k \prec U_i$, then the complex is equivalent to a ranking; to each of the objects U_i may be assigned an integer u_i giving the position of that object in the ranking. Thus in the case $n = 4$, the comparisons

$$U_1 \succ U_2, \quad U_1 \succ U_3, \quad U_1 \succ U_4, \quad U_2 \prec U_3, \quad U_2 \prec U_4, \quad U_3 \succ U_4$$

give the ranking

$$U_2 \prec U_4 \prec U_3 \prec U_1,$$

which may be expressed by

$$u_1 = 4, \quad u_2 = 1, \quad u_3 = 3, \quad u_4 = 2.$$

We then have

$$u_2 < u_4 < u_3 < u_1.$$

(iii) If the complex of comparisons is inconsistent, no ranking is possible. The procedure (i) is then repeated as often as necessary until a consistent complex is obtained; and the corresponding ranking is accepted.

The above is a possible way of generating rankings. It should be emphasized that it is not suggested that any actual experiment would be performed in this fashion; if the paired-comparisons *can* be made independently (and recorded), there is little point in trying to force the preferences into a ranking. What is here attempted is the production of a model for situations where the observed data is a ranking, the difficulty being that the comparisons are no longer independent (thus if $U_i \prec U_j$ and $U_j \prec U_k$, we must have $U_i \prec U_k$). We are introducing this dependence by considering a conditional distribution of paired-comparisons; we admit only a subset (containing $n!$ members) of the total $2^{{}^nC_2}$ possible outcomes.

The probability of obtaining any given ranking (u) from the above construction will be proportional to the probability under the Babington Smith model of the corresponding complex of comparisons. Thus the probability of the ranking

$$(u) \equiv (u_1, u_2, u_3, u_4) = (4, 1, 3, 2)$$

above is proportional to

$$(1-\pi_{12})(1-\pi_{13})(1-\pi_{14})\pi_{23}\pi_{24}(1-\pi_{34});$$

putting

$$\lambda_{ij}^2 = \pi_{ij}/(1-\pi_{ij}),$$

we have that the probability of the general ranking $(u) = (u_1, u_2, \ldots, u_n)$ under this hypothesis is

$$\mathscr{H}_\lambda: \quad P\{(u)\} = K_n \prod_{1 \leqslant i < j \leqslant n} \lambda_{ij}^{\operatorname{sgn}(u_j - u_i)}, \tag{1}$$

where K_n is so chosen that the probabilities sum to 1; i.e.

$$K_n^{-1} = \sum_{(u)} \prod_{1 \leqslant i < j \leqslant n} \lambda_{ij}^{\operatorname{sgn}(u_j - u_i)} = P\{\text{comparisons are consistent}\} \prod_{1 \leqslant i < j \leqslant n} [\pi_{ij}(1-\pi_{ij})]^{-\frac{1}{2}}.$$

3. Specializations

The above general model is rather cumbersome; we now consider possible ways of specializing it so as to reduce the number of parameters. We shall attempt to find models which simplify

(i) the probabilities $P\{(u)\}$ of the various rankings (u);

(ii) the dist ibution problems connected with the two coefficients r_s and t_k;

(iii) various other interesting and important quantities, such as

$$p_{ij} = P\{u_i < u_j \mid \text{all comparisons consistent}\}, \tag{2}$$

which are complicated in the general case.

4. The Bradley-Terry model

In the paired-comparison case, Bradley & Terry (1952) reduce the number of parameters from ${}^nC_2$ to (effectively) $n-1$ by assuming that

$$\pi_{ij} = P\{U_i \prec U_j\} = \pi_j/(\pi_i + \pi_j) \quad (1 \leqslant i < j \leqslant n)$$

for some non-negative numbers $\pi_1, \pi_2, \ldots, \pi_n$. This gives

$$\lambda_{ij}^2 = \pi_j/\pi_i, \tag{3}$$

and for the proposed ranking model, we have

$$\mathscr{H}_B: \quad P\{(u)\} \propto \prod_{1 \leqslant i < j \leqslant n} (\pi_j/\pi_i)^{\frac{1}{2}\operatorname{sgn}(u_j - u_i)} = \prod_{1 \leqslant i \leqslant n} \pi_i^{\frac{1}{2}\sum_j \operatorname{sgn}(u_i - u_j)} \propto \prod_{1 \leqslant i \leqslant n} \pi_i^{u_i}. \tag{4}$$

This is a term in the expansion of the permanent

$$\begin{vmatrix} + & & & + \\ \pi_1 & \pi_1^2 & \ldots & \pi_1^n \\ \pi_2 & \pi_2^2 & \ldots & \pi_2^n \\ \vdots & \vdots & & \vdots \\ \pi_n & \pi_n^2 & \ldots & \pi_n^n \end{vmatrix} = \Big| \pi_i^j \Big|_{i,j=1,2,\ldots,n}^{+}. \tag{5}$$

We note that this model is closely associated with what we may call a 'generalized matching coefficient', this being taken to mean a coefficient

$$A(u) = \sum_{1 \leqslant i \leqslant n} \alpha_{iu_i},$$

where the $\{\alpha_{ij}\}$ can take any values whatsoever. Examples of measures which are essentially of this form are:

(i) Number of matches; put $\alpha_{ij} = \delta_i^j$ (Kronecker δ).

(ii) Spearman's r_s; put $\alpha_{ij} = (i-j)^2$.

(iii) Spearman's 'footrule'; put $\alpha_{ij} = |i-j|$.

Others which may be considered are:

(iv) Number of matches + near misses; put $\alpha_{ij} = \delta_i^{j-1} + \delta_i^j + \delta_i^{j+1}$.

(v) Weighted form of (iv); put $\alpha_{ij} = \delta_i^{j-1} + K\delta_i^j + \delta_i^{j+1}$, where, for example, $K = 2$.

In the above cases, the various measures, $C(u)$ say, are equal either to the corresponding $A(u)$ directly, or to a simple function of $A(u)$; we have

$$C(u) = F(A(u)) \quad \text{for all } (u), \tag{6}$$

where also

$$A(u) = F^{-1}(C(u)) \quad \text{for all } (u). \tag{7}$$

The score $S(u)$ for Kendall's t_k (see eqn. (8)) can be expressed in the form (6), but for $n \geqslant 4$ the inverse function as in (7) does not exist (see Appendix I).

The p.g.f. of $A(u)$ for the above model is simply

$$\mathscr{H}_B: \quad \sum_{(u)} P\{(u)\} z^{A(u)} = \left|\overset{+}{z^{a_{ij}}}\right| \Big/ \left|\overset{+}{\pi_i^j}\right|.$$

However, in the absence of simple methods of manipulating permanents, we shall not investigate this model further at present.

Barton & David (1956) have proposed an alternative hypothesis based on distorting the null matching distribution.

5. An alternative method

We now consider an alternative method of simplification of the model $\mathscr{H}_\lambda$. The most general non-null model possible would specify the probability of each ranking separately, needing $n! - 1$ parameters. We have reduced this number to ${}^nC_2$ by assuming a special structure for the model; it can only be further reduced by even more restrictive assumptions. Consider the following additional assumption:

Assumption A. Rankings (u), giving (when compared with the standard ranking $(1, 2, \ldots, n)$) the same values for both r_s and t_k, have the same probability.

This assumption is equivalent to

Assumption A'. The probability of any ranking (u) is a function only of

$$\left.\begin{aligned} R(u) &= \sum_{1 \leqslant i < j \leqslant n} (j-i)\operatorname{sgn}(u_j - u_i) \\ &= 2\sum_{1 \leqslant i \leqslant n} iu_i - \tfrac{1}{2}n(n+1)^2 \\ &= \tfrac{1}{6}n(n^2-1)r_s \\ \text{and} \quad S(u) &= \sum_{1 \leqslant i < j \leqslant n} \operatorname{sgn}(u_j - u_i) \\ &= \tfrac{1}{2}n(n-1)t_k. \end{aligned}\right\} \tag{8}$$

Consider also the following assumption:

Assumption B. Pairs of rankings which are inverses of one another have the same probability.

Two rankings (u), (v) are said to be inverses (sometimes 'conjugates') if when $u_i = j$, then $v_j = i$. Thus the rankings $(u) = (3, 5, 2, 4, 1)$ and $(v) = (5, 3, 1, 4, 2)$ are inverses since

$$u_1 = 3,\ v_3 = 1; \quad u_2 = 5,\ v_5 = 2; \quad \text{etc.}$$

It is well known (see, for example, Kendall, 1955, p. 6 (*not* p. 11)) that for such a pair of rankings

$$R(u) = R(v), \quad S(u) = S(v),$$

and so Assumption B is contained in Assumption A.

From Assumption B and eqn. (1) we may obtain certain relations between the parameters λ_{ij}; thus for the pair of rankings above we obtain

$$P\{(u)\} = K_5\lambda_{12}\lambda_{13}^{-1}\lambda_{14}\lambda_{15}^{-1}\lambda_{23}^{-1}\lambda_{24}^{-1}\lambda_{25}^{-1}\lambda_{34}\lambda_{35}^{-1}\lambda_{45}^{-1},$$
$$P\{(v)\} = K_5\lambda_{12}^{-1}\lambda_{13}^{-1}\lambda_{14}^{-1}\lambda_{15}^{-1}\lambda_{23}^{-1}\lambda_{24}\lambda_{25}^{-1}\lambda_{34}\lambda_{35}\lambda_{45}^{-1}$$

and for these to be equal we must have

$$\lambda_{12}^2\lambda_{14}^2 = \lambda_{24}^2\lambda_{35}^2.$$

Proceeding thus for all pairs of inverses we obtain the following relations (for $n = 5$):

$$\lambda_{12} = \lambda_{23} = \lambda_{34} = \lambda_{45} = \lambda_1, \quad \text{say}$$
$$\lambda_{13} = \lambda_{24} = \lambda_{35} = \lambda_2, \quad \text{say}$$
$$\lambda_{14} = \lambda_{25} = \lambda_3, \quad \text{say}$$

and we put $\lambda_{15} = \lambda_4$.

Further, $\lambda_1\lambda_3 = \lambda_2^2, \quad \lambda_2\lambda_4 = \lambda_3^2.$

Similar relations (in fact subsets of the above) are obtained for $n = 3, 4$. These relations are generalized in Appendix II and show that Assumption B contains

Assumption C. For any (fixed) n, we have

(i) $\lambda_{ij} = \lambda_{j-i} \quad (1 \leqslant i < j \leqslant n)$;

(ii) $\lambda_{k-1}\lambda_{k+1} = \lambda_k^2 \quad (2 \leqslant k \leqslant n-1)$.

Assumption C (ii) implies that the $\{\lambda_k\}$ are in geometrical progression, i.e. for some θ, ϕ we have

$$\lambda_k = \theta^k\phi,$$

i.e. $$P\{U_i \prec U_{i+k}\} = \frac{\theta^k\phi}{\theta^k\phi + \theta^{-k}\phi^{-1}} = \tfrac{1}{2} + \tfrac{1}{2}\tanh(k\log\theta + \log\phi).$$

Hence $$\mathscr{H}(\theta, \phi): \quad P\{(u)\} = K_n \prod_{1 \leqslant i < j \leqslant n} (\theta^{j-i}\phi)^{\operatorname{sgn}(u_j - u_i)}$$
$$= K_n\theta^{R(u)}\phi^{S(u)}. \tag{9}$$

Assumption C thus contains Assumption A. We deduce that all three assumptions are equivalent. We shall take (9) to be our basic non-null model, $\mathscr{H}(\theta, \phi)$.

6. Properties of the model $\mathscr{H}(\theta, \phi)$

Certain properties of this model are immediate.

(*a*) Let the null-hypothesis probability of a pair of values (R, S) be

$$P_0(R, S) = (n!)^{-1}F(R, S),$$

so that there are just $F(R,S)$ different rankings (u) having $R(u)=R$, $S(u)=S$. Then for the present model we have

$$\mathscr{H}(\theta,\phi): \quad P_{\theta,\phi}(R,S) \propto \theta^R \phi^S P_0(R,S).$$

If the p.g.f. under the null hypothesis is

$$\mathscr{H}_0: \quad M_0(x,y) = \sum_{R,S} P_0(R,S)\, x^R y^S,$$

then the p.g.f. for the present model is

$$\mathscr{H}(\theta,\phi): \quad M_{\theta,\phi}(x,y) = M_0(x\theta, y\phi)/M_0(\theta,\phi). \tag{10}$$

(*b*) The present model will be a special case of the one considered in § 4 if (from eqn. (3))

$$\pi_{i+k} = \theta^{2k}\phi^2 \pi_i \quad (1 \leqslant i < i+k \leqslant n);$$

this requires $\phi = 1$, and then we may take without loss of generality

$$\pi_i = \theta^{2i} \quad (1 \leqslant i \leqslant n).$$

Then
$$P\{U_i \prec U_j\} = \theta^k/(\theta^k + \theta^{-k})$$

and
$$\mathscr{H}(\theta): \quad P\{(u)\} = \theta^{2\Sigma i u_i} \Big/ \overset{+}{\Big|} \theta^{2ij} \overset{+}{\Big|} \propto \theta^{R(u)}. \tag{11}$$

As with $\mathscr{H}_B$, this model is unmanageable because of the permanents involved.

(*c*) It is known (Daniels, 1944; Hoeffding, 1948) that under the null hypothesis, the joint distribution of R and S tends to the bivariate Normal form; i.e. that for any suitable† set $\mathscr{A}$ of points (R,S) containing α such points we have

$$\sum_{(R,S)\in\mathscr{A}} P_0(R,S) \simeq \iint_{\mathscr{A}^*} f(R,S)\, dR\, dS,$$

where $\mathscr{A}^*$ is a convex region of area 4α containing $\mathscr{A}$, but no points of $\bar{\mathscr{A}}$; the factor 4 enters since $\Delta R = \Delta S = 2$. $f(R,S)$ is the bivariate Normal distribution function with

$$\mu_R = 0, \quad \sigma_R^2 = \tfrac{1}{36} n^2 (n-1)(n+1)^2,$$
$$\mu_S = 0, \quad \sigma_S^2 = \tfrac{1}{18} n (n-1)(2n+5), \qquad \rho = \frac{2(n+1)}{\sqrt{\{2n(2n+5)\}}} \simeq 1 - \frac{1}{4n}.$$

For the model $\mathscr{H}(\theta,\phi)$ we have approximately (for $\mathscr{A}$ 'small')

$$\sum_{(R,S)\in\mathscr{A}} P_{\theta,\phi}(R,S) \simeq \theta^R \phi^S \iint_{\mathscr{A}^*} f(R,S)\, dR\, dS,$$

whence asymptotically, under suitable approaches to the limit for θ and ϕ, R and S are bivariate Normal with

$$\mathscr{E}(R) \simeq \sigma_R(\sigma_R \log\theta + \rho\sigma_S \log\phi), \quad \operatorname{var} R \simeq \sigma_R^2,$$
$$\mathscr{E}(S) \simeq \sigma_S(\rho\sigma_R \log\theta + \sigma_S \log\phi), \quad \operatorname{var} S \simeq \sigma_S^2, \qquad \rho(R,S) \simeq \rho.$$

† The usual proofs demonstrate only that the *cumulative* distribution tends to the Normal form; what is required here is that the *ordinates*, averaged over α points, tend to the Normal ordinates. I believe that the result for $\alpha = 1$ has not been proved; Haden (1947) has proved the result for the marginal distribution of S, but in view of the well-known erratic behaviour of R the bivariate result is more difficult. I conjecture it to be true; however, for the present we only need to consider sets with

$$\alpha = O(n^{-1}\sigma_R\sigma_S) = O(n^3).$$

Conditions under which this is the correct limiting form are as follows. The asymptotic result in the null case may be written

$$\lim_{n\to\infty}\left\{\sum_{\substack{R\leqslant R_0\\ S\leqslant S_0}} P_0(R,S)\Big/\int_{-\infty}^{R_0}\int_{-\infty}^{S_0} f(R,S)\,dR\,dS\right\}=1,$$

where, as $n\to\infty$,

$$R_0/\sigma_R = O(1),\quad S_0/\sigma_S = O(1),$$
$$R_0/\sigma_R - S_0/\sigma_S = O(n^{-1}).$$

This last condition follows from the approach of ρ to 1. (It is possible that weaker conditions (e.g. $R_0/\sigma_R = O(n^{\frac{1}{2}})$) may suffice; however, this is not known.) For the non-null asymptotic result to hold

$$\mathscr{E}(R)/\sigma_R = O(1),\quad \mathscr{E}(S)/\sigma_S = O(1),$$
$$\mathscr{E}(R)/\sigma_R - \mathscr{E}(S)/\sigma_S = O(n^{-1}),$$

whence
$$\sigma_R\log\theta = O(1),\quad \sigma_S\log\phi = O(1),$$

i.e.
$$\log\theta = O(n^{-\frac{1}{2}}),\quad \log\phi = O(n^{-\frac{1}{2}}).$$

Putting
$$\lim\sigma_R\log\theta = \mu',\quad \lim\sigma_S\log\phi = \mu'',$$

we have
$$\lim\mathscr{E}(R)/\sigma_R = \mu'+\mu'' = \lim\mathscr{E}(S)/\sigma_S.$$

Thus the two parameters are asymptotically indistinguishable; each of them merely shifts the bivariate distribution in the direction $R/\sigma_R = S/\sigma_S$. We deduce further that R and S are asymptotically equivalent in the model $\mathscr{H}(\theta,\phi)$, i.e. they have asymptotically equal power for detecting a change in $(\sigma_R\log\theta+\sigma_S\log\phi)$, at least whenever the above is the correct limiting form.

(*d*) From the form of eqn. (9), $R(u)$ and $S(u)$ are respectively (and jointly) sufficient for θ and ϕ; they will therefore provide most-efficient estimators whenever their asymptotic joint distribution is Normal. This brings out a distinction between the present model (which is a pure ranking model) and other models which have been suggested. Thus if a sample of n is drawn from a bivariate Normal population with correlation ρ, and the variate values are replaced by ranks, then Pitman's asymptotic relative efficiency (A.R.E.) of R (or S) in estimating ρ (relative to the product-moment correlation) is known to be $9\pi^{-2}$ (Hotelling & Pabst, 1936; Moran, 1951). Lehmann (1953) considers a certain non-parametric system of alternatives to independence, and shows that R gives the optimum test for these alternatives, for all sample sizes. For the Thurstone–Babington Smith model, where a judge arrives at a ranking of n objects by first making an estimate of the 'value' of each object, these estimates being independently and Normally distributed about a regression depending on the true ranking of the objects (for simplicity the means for the different objects are usually taken to be in arithmetic progression), Stuart (1954) has shown that the A.R.E. of R or S (relative to the product-moment regression coefficient) is $3\pi^{-1}$. It is hoped in a further communication to investigate the relations between the present models and these alternatives.

7. The model $\mathscr{H}(\phi)$

We shall now concentrate on the parameter ϕ. Putting $\theta = 1$ in the basic model (9) gives the 'ϕ-model':

$$\mathscr{H}(\phi):\quad P\{(u)\} = K_n\phi^{S(u)} \tag{12}$$

and
$$P\{U_i \prec U_j\} = \pi = \phi/(\phi+\phi^{-1})\quad (1\leqslant i<j\leqslant n).$$

Writing now (u_n) for (u) to denote the dependence on n, the p.g.f. of $S_n = S(u_n)$ under the null hypothesis is known to be (see, for example, Moran, 1950)

$$M_n(z) = \sum_{(u_n)} P_0\{(u_n)\}\, z^{S(u_n)} = \prod_{1\leqslant k\leqslant n} \frac{1}{k}(z^{k-1}+z^{k-3}+\ldots+z^{1-k}), \tag{13}$$

whence the random variable S_n can be expressed as the sum of n independent random variables s_k with p.g.f.'s

$$m_k(z) = \frac{1}{k}\frac{z^k - z^{-k}}{z - z^{-1}} = \frac{\sin kx}{k \sin x},$$

where $z = e^{ix}$. Thus s_k has a uniform distribution on the k points $s_k = k-1, k-3, \ldots, 1-k$. Hence Moran (1950) obtains the cumulant generating function (c.g.f.) of s_k as

$$\log m_k(e^{ix}) = \sum_{m=1}^{\infty} \frac{(ix)^{2m}}{2m!}(-1)^m \frac{B_m 2^{2m-1}}{m}(1-k^{2m}). \tag{14}$$

For the present model we have

$$\begin{aligned} M_{n,\phi}(z) &= M_n(\phi z)/M_n(\phi) \\ &= \prod_{1\leqslant k\leqslant n} \frac{\phi^k z^k - \phi^{-k} z^{-k}}{\phi z - \phi^{-1} z^{-1}} \frac{\phi - \phi^{-1}}{\phi^k - \phi^{-k}}, \end{aligned} \tag{15}$$

whence S_n can again be expressed as the sum of n independent random variables s_k, with p.g.f.'s

$$m_{k,\phi}(z) = m_k(\phi z)/m_k(\phi).$$

Thus s_k now has a distribution on the k points $k-1, k-3, \ldots, 1-k$ with probabilities in geometric progression, proportional to ϕ^{s_k}. Also

$$K_n^{-1} = M_n(\phi). \tag{16}$$

Putting $\log \phi = \delta$ we have for the c.g.f. of s_k

$$\log m_{k,\phi}(e^{ix}) = \log \frac{\sinh(kix + k\delta)\sinh\delta}{\sinh(ix+\delta)\sinh k\delta};$$

hence, writing

$$\beta_k = \coth k\delta = (\phi^k + \phi^{-k})(\phi^k - \phi^{-k})^{-1},$$

we have

$$\begin{aligned} \kappa_{1,k} &= k\beta_k - \beta_1, \\ \kappa_{2,k} &= (\beta_1^2 - 1) - k^2(\beta_k^2 - 1), \\ \kappa_{3,k} &= 2k^3\beta_k(\beta_k^2 - 1) - 2\beta_1(\beta_1^2 - 1), \\ \kappa_{4,k} &= 2(\beta_1^2 - 1)(3\beta_1^2 - 1) - 2k^4(\beta_k^2 - 1)(3\beta_k^2 - 1). \end{aligned}$$

These reduce to Moran's values when $\delta \to 0$. An extreme case is obtained when $k \to \infty$ with ϕ constant $\neq 1$; $\beta_k \sim 1 + 2\phi^{-2k}$ and we find

$$\begin{aligned} \kappa_{1,k} &\sim k - \beta_1 + 2k\phi^{-2k}, \\ \kappa_{2,k} &\sim (\beta_1^2 - 1) - 4k^2\phi^{-2k}, \\ \gamma_1 &\to -2\beta_1(\beta_1^2 - 1)^{-\frac{1}{2}} = -(\phi + \phi^{-1}), \\ \gamma_2 &\to 2(3\beta_1^2 - 1)(\beta_1^2 - 1)^{-1} = 2 + \gamma_1^2. \end{aligned}$$

The corresponding limiting distribution of s_k is geometric;

$$P\{s_k = k - 1 - 2j\} = (1 - \phi^{-2})\phi^{-2j} \quad (j = 0, 1, 2, \ldots). \tag{17}$$

Thus even in this case, the central limit theorem will apply; asymptotically,

$$(S_n - \tfrac{1}{2}n(n+1) + n\beta_1)\,(n(\beta_1^2 - 1))^{-\frac{1}{2}}$$

is a unit Normal variable. We notice that this is essentially a different limit from that considered in § 6; it is not known whether the corresponding limits for $R(u_n)$ and for the bivariate distribution are Normal, or what happens to the correlation between R_n and S_n. In intermediate cases, when $\log \phi \to 0$ slower than $n^{-\frac{1}{2}}$, we would expect the limiting distribution of S_n to be Normal, with $\mathscr{E}(S_n) = o(n^2)$, $\operatorname{var} S_n = o(n)$.

8. Decomposition of s_k

In the null case, Feller (1945) has shown that the variables s_k may be taken as

$$s_k = \sum_{1 \leqslant i \leqslant k-1} \operatorname{sgn}(u_k - u_i);$$

this representation may be carried over to the ϕ-model case. (This contradicts a remark of Moran's (1950).) Let us go back to the original paired-comparison approach. Suppose the judge has decided the ${}^{n-1}C_2$ preferences among $U_1, U_2, \ldots, U_{n-1}$ and has as yet no inconsistencies. Then we have an $(n-1)$-ranking (u_{n-1}) say; and the p.g.f. of $S(u_{n-1})$ is $M_{n-1,\phi}(z)$. Now the judge makes the final set of $n-1$ comparisons, $U_i \prec$ or $\succ U_n$ for $i = 1, 2, \ldots, n-1$. These comparisons are independent amongst themselves and of the preferences already decided:

$$P\{U_i \prec U_n\} = \pi \quad (1 \leqslant i \leqslant n-1).$$

Of the 2^{n-1} possible sets of outcomes for these comparisons, only these sets will be consistent which insert U_n into one of the n intervals between the $n-1$ U's already ranked; the probability that U_n is ranked higher than the lower $j-1$ U's and lower than the higher $n-j$ U's is

$$\pi^{j-1}(1-\pi)^{n-j} \quad (1 \leqslant j \leqslant n)$$

independently of the ranking (u_{n-1}).

This set of comparisons gives a ranking (u_n) with

$$S(u_n) = S(u_{n-1}) + 2j - 1 - n = S(u_{n-1}) + s_n,$$

say. The probability that the comparisons are consistent is thus

$$\sum_{1 \leqslant j \leqslant n} \pi^{j-1}(1-\pi)^{n-j} = (\pi^n - (1-\pi)^n)\,(2\pi - 1)^{-1} = C_n \quad \text{say};$$

and we have as before that s_n is a random variable independent of $S(u_{n-1})$ (and of (u_{n-1})) with

$$P\{s_n\} = C_n^{-1}\pi^{\frac{1}{2}(n-1+s_n)}(1-\pi)^{\frac{1}{2}(n-1-s_n)} = \frac{\phi - \phi^{-1}}{\phi^n - \phi^{-n}}\phi^{s_n} \quad (s_n = 2j-1-n;\ 1 \leqslant j \leqslant n). \tag{18}$$

9. An invariance property

We now prove for the ϕ-model an important invariance property, namely, that

$$p_{ij}^{(n)} = P\{u_i < u_j \mid n, \phi, \text{comparisons are consistent}\}$$

depends only on $j-i$ and ϕ, and not on i or n. We prove that

$$p_{i\,i+m-1}^{(n)} = p_{1m}^{(m)} \quad \text{for} \quad 1 \leqslant i < i+m-1 \leqslant n.$$

This will be proved by induction on n. It is true by definition for $n = m$; assume that it has been proved for $n = m, m+1, \ldots, a$. Fix i, with $1 \leqslant i \leqslant a-m+1$. Consider the rankings (u_a) with $S(u_a) = S$ and having the property (property L) that $u_i < u_{i+m-1}$; we put $\sum_{(u_a)|L,S}$ to denote summation over these rankings. We shall write $((u_a)j)$ for a ranking (u_{a+1}) which has $u_{a+1} = j$. We have by the inductive hypothesis

$$p_{1m}^{(m)} = p_{ii+m-1}^{(a)} = \sum_S \sum_{(u_a)|L,S} P\{(u_a)\}.$$

Also
$$p_{ii+m-1}^{(a+1)} = \sum_S \sum_{(u_{a+1})|L,S} P\{(u_{a+1})\} = \sum_S \sum_{1 \leqslant j \leqslant a+1} \sum_{\substack{(u_{a+1})|L,S \\ u_{a+1}=j}} P\{((u_a)j)\}.$$

Now $S((u_a)j) = S(u_a) + 2j - 2 - a$, and $((u_a)j)$ has the property L if and only if (u_a) has; hence

$$\begin{aligned} p_{ii+m-1}^{(a+1)} &= \sum_S \sum_{1 \leqslant j \leqslant a+1} \sum_{(u_a)|L,S} P\{((u_a)j)\} \\ &= K_{a+1} \sum_{1 \leqslant j \leqslant a+1} \phi^{2j-2-a} \sum_S \sum_{(u_a)|L,S} \phi^S \\ &= K_{a+1} \frac{\phi^{a+1} - \phi^{-a-1}}{\phi - \phi^{-1}} K_a^{-1} p_{1m}^{(m)} = p_{1m}^{(m)}. \end{aligned}$$

This proves the result for all i except $i = (a+1) - m + 1$; this case can be proved similarly by considering rankings $(u_{a+1}) = (j(u_a))$. Hence by induction on a we have finally $p_{ii+k}^{(n)} = p_k$ $(1 \leqslant i < i+k \leqslant n)$, where the $\{p_k\}$ depend only on ϕ.

We may obtain the $\{p_k\}$ explicitly as follows: we write $(i(u_{k-2})j)$ for a ranking (u_k) which has $u_1 = i$, $u_k = j$. Then

$$S(i(u_{k-2})j) = S(u_{k-2}) + 2(j-i) + \operatorname{sgn}(i-j)$$

and
$$\begin{aligned} p_{k-1} = p_{1k}^{(k)} &= \sum_{1 \leqslant i < j \leqslant k} \sum_{(u_{k-2})} P\{(i(u_{k-2})j)\} \\ &= K_k \sum_{1 \leqslant i < j \leqslant k} \phi^{2(j-i)-1} \sum_{(u_{k-2})} \phi^{S(u_{k-2})} \\ &= K_k (\phi^{2k-1} - k\phi + (k-1)\phi^{-1})(\phi - \phi^{-1})^{-2} K_{k-2}^{-1}, \end{aligned}$$

whence for any i $(1 \leqslant i < i+k \leqslant n)$

$$\begin{aligned} \gamma_k &= \mathscr{E}(\operatorname{sgn}(u_{i+k} - u_i)) = 2p_k - 1 \\ &= (k+1)\frac{\phi^{k+1} + \phi^{-k-1}}{\phi^{k+1} - \phi^{-k-1}} - k\frac{\phi^k + \phi^{-k}}{\phi^k - \phi^{-k}} = (k+1)\beta_{k+1} - k\beta_k \end{aligned} \tag{19}$$

in the notation of § 7.

From this result we have

$$\begin{aligned} \mathscr{E}(S_n) = \mathscr{E} \sum_{1 \leqslant i < j \leqslant n} \operatorname{sgn}(u_j - u_i) &= \sum_{1 \leqslant i < j \leqslant n} \gamma_{j-i} = \sum_{1 \leqslant k \leqslant n-1} (n-k)\gamma_k, \\ \mathscr{E}(S_n) - \mathscr{E}(S_{n-1}) &= \sum_{1 \leqslant k \leqslant n-1} \gamma_k = k\beta_k - \beta_1, \end{aligned}$$

in agreement with § 7. Similarly we may obtain $\mathscr{E}(R_n)$:

$$\begin{aligned} \mathscr{E}(R_n) = \mathscr{E} \sum_{1 \leqslant i < j \leqslant n} (j-i)\operatorname{sgn}(u_j - u_i) &= \sum_{1 \leqslant i < j \leqslant n} (j-i)\gamma_{j-i} \\ &= \sum_{1 \leqslant k \leqslant n-1} k(n-k)\gamma_k = \sum_{1 \leqslant k \leqslant n} k(2k-n-1)\beta_k \\ &= \sum_{1 \leqslant k \leqslant n} (2k-n-1)(k\beta_k - \beta_1). \end{aligned} \tag{20}$$

10. Interpretation of the parameters

Having now considered some of the consequences of the Assumptions of § 5, we can give interpretations of the two parameters θ and ϕ. Each of them corresponds to a certain type of departure from the null hypothesis, that the judge cannot detect the real differences between the objects:

(*a*) The model $\mathscr{H}(\theta)$, by its association with the Bradley–Terry model, may be considered to be assigning a weight $w(i) = \theta^{2i}$ to each of the objects U_i to be ranked; this weight remains constant throughout the process of making the paired-comparisons and arriving at the ranking. The paired-comparisons are made according to the rule

$$P\{U_i \prec U_j\} = \frac{w(j)}{w(i)+w(j)} = \frac{\theta^{2j}}{\theta^{2i}+\theta^{2j}} = \frac{\theta^k}{\theta^k+\theta^{-k}},$$

where $k = j-i$ for $1 \leqslant i < j \leqslant n$. Thus this probability depends on the difference between the *a priori* ranks of the objects, but not on their absolute positions. While this seems an attractive hypothesis, it must be remembered that the rejection of inconsistent sets of comparisons introduces a distortion, the effect of which it is difficult to assess.

Table 1. $\mathscr{H}(\phi)$: *Values of* $p_k = P\{u_i < u_{i+k}\}$

k \ ϕ^2 / π	1·2 0·5̇4̇	1·5 0·6	2·0 0·6̂
1	0·5455	0·6000	0·6667
2	0·5754	0·6632	0·7619
3	0·6049	0·7215	0·8386
4	0·6337	0·7737	0·8946
5	0·6617	0·8191	0·9339
6	0·6887	0·8577	0·9599
7	0·7145	0·8897	0·9763
8	0·7392	0·9155	0·9862
9	0·7626	0·9361	0·9922

(*b*) The model $\mathscr{H}(\phi)$ puts for the paired-comparisons

$$P\{U_i \prec U_j\} = \pi = \frac{\phi}{\phi+\phi^{-1}} \quad (1 \leqslant i < j \leqslant n), \tag{21}$$

this probability being independent of both i and j. This may not seem as reasonable an assumption as that for the θ-model above; however, the quantities we are really interested in are the p_{ij} of eqn. (2).

Table 1 gives the values of

$$p_k = P\{u_i < u_{i+k}\} \quad (1 \leqslant i < i+k \leqslant n)$$

(from (19)) for three values of ϕ^2, namely, 1·2, 1·5 and 2·0 corresponding to

$$\pi = \tfrac{6}{11} = 0{\cdot}5\dot{4}, 0{\cdot}6, \tfrac{2}{3}.$$

It was proved in § 9 that p_k is independent of i and n.

It will be seen that these values progress in a very reasonable manner—the further apart in the true ranking two objects are, the more probably will the judge put them in the correct order.

From eqn. (18) we have a further interpretation of the parameter ϕ. Given that the objects $U_1, U_2, \ldots, U_{n-1}$ have been compared and a ranking obtained, the judge now compares U_n with each of these; each comparison being independent of the rest, and according to (21). Now the restriction that only consistent comparisons are accepted means (eqn. (18)) that U_n takes the jth position in the ranking with probability proportional to ϕ^{2j}, i.e. (if $\phi > 1$) the probability is largest that U_n take its correct position ($j = n$), and falls off geometrically as j decreases to 1.

(c) The general model $\mathscr{H}(\theta, \phi)$ (eqn. (9)) puts

$$P\{U_i \prec U_j\} = \frac{\theta^k \phi}{\theta^k \phi + \theta^{-k} \phi^{-1}} = \tfrac{1}{2} + \tfrac{1}{2} \tanh (k \log \theta + \log \phi)$$

for $k = j - 1$, $1 \leqslant i < j \leqslant n$. This is a simple monotonic function of $k \log \theta + \log \phi$; thus the effects of θ and ϕ are in a sense additive. The term $\log \phi$ gives a departure from the null hypothesis which is the same for all pairs (i, j); the other term adds to this a departure depending linearly on $k = j - i$. This model would appear to be an attractive approximation to the general model $\mathscr{H}_\lambda$ (eqn. (1)).

11. Applications of the ϕ-model

From the form of the model $\mathscr{H}(\phi)$ (eqn. (12)) we have that $S(u)$ is sufficient for ϕ; problems of estimation and of hypothesis-testing will therefore be most naturally expressed in terms of S. However, since R and S are asymptotically equivalent, any procedure using S may if desired be replaced (approximately) by one using R.

We shall use the following notation:

$$\delta = \log \phi, \quad \mathscr{E}_\phi = \mathscr{E}(S \mid \mathscr{H}(\phi)), \quad V_\phi = \operatorname{var}(S \mid \mathscr{H}(\phi)),$$

$$\Delta = \mathscr{E}_\phi V_\phi^{-\frac{1}{2}}, \quad V_0 = \operatorname{var}(S \mid \mathscr{H}_0) = \tfrac{1}{18} n(n-1)(2n+5),$$

which though not entirely consistent, will not lead to confusion.

Approximations. From Moran's expression (14) we have an expansion of $\log M_n(\phi)$ (eqn. (13)) in powers of δ:

$$\log M_n(\phi) = \frac{\delta^2}{6} \sum_{k=2}^{n} (k^2 - 1) - \frac{\delta^4}{180} \sum_{k=2}^{n} (k^4 - 1) + \ldots$$

$$= \tfrac{1}{2} V_0 \delta^2 - \frac{n(n-1)}{5400} (6n^3 + 21n^2 + 31n + 31) \delta^4 + \ldots$$

whence from (15)

$$\mathscr{E}_\phi = \frac{\partial}{\partial z} \frac{M_n(\phi z)}{M_n(\phi)}\bigg|_{z=1} = \phi \frac{d}{d\phi} \log M_n(\phi) = V_0 \delta - \frac{n(n-1)}{1350} (6n^3 + 21n^2 + 31n + 31) \delta^3 + \ldots,$$

$$V_\phi = \phi \frac{d}{d\phi} \log M_n(\phi) + \phi^2 \frac{d^2}{d\phi^2} \log M_n(\phi) = V_0 \left(1 - \frac{\delta^2}{25} \frac{6n^3 + 21n^2 + 31n + 31}{2n + 5} + \ldots\right).$$

We notice the curious result

$$V_\phi = \phi \frac{d}{d\phi} \mathscr{E}_\phi.$$

We obtain further

$$V_\phi = V_0\left(1 - \frac{27}{25}\frac{\Delta^2}{n} + \left(\frac{27}{50} + \frac{9936}{30625}\Delta^2\right)\frac{\Delta^2}{n^2} + \ldots\right).$$

Hence

$$V_\phi \simeq V'_\phi = V_0 - \frac{27}{25}\frac{\mathscr{E}_\phi^2}{n}.$$

Table 2 gives values of the percentage error of the approximation V'_ϕ for various values of n and ϕ, with the corresponding values of Δ. It will be seen that the approximation is remarkably good even for small n, provided Δ is not too large; except for very small n, the approximation improves with increasing n if Δ is held constant (compare the three entries with $\Delta = 1{\cdot}72$ or $1{\cdot}73$).

Table 2. *$\mathscr{H}(\phi)$: Percentage error of the approximation V'_ϕ, with corresponding values of $\Delta = \mathscr{E}_\phi V_\phi^{-\frac{1}{2}}$*

$$\% \text{ error} = 100(V'_\phi - V_\phi)\, V_\phi^{-1}$$

n	$\phi^2 = 1{\cdot}1$		$\phi^2 = 1{\cdot}2$		$\phi^2 = 1{\cdot}5$		$\phi^2 = 2{\cdot}0$	
	% error	Δ	% error	Δ	% error	Δ	% error	Δ
2	+0·11	0·05	+0·38	0·09	+1·92	0·20	+ 5·75	0·35
4	+0·08	0·14	+0·31	0·27	+1·31	0·61	+ 3·15	1·07
6	+0·05	0·25	+0·17	0·49	+0·08	1·12	− 4·09	2·01
8	+0·02	0·39	−0·02	0·75	−2·35	1·73	−19·28	3·17
10	−0·02	0·54	−0·30	1·04	−6·63	2·44		
12	−0·07	0·70	−0·71	1·36				
14	−0·14	0·88	−1·32	1·72				
16	−0·23	1·07	−2·16	2·11				
18	−0·34	1·28						
20	−0·49	1·50						
22	−0·67	1·73						
24	−0·90	1·97						
26	−1·18	2·23						

Under $\mathscr{H}(\phi)$, S is approximately Normally distributed with mean $\mathscr{E}_\phi$ and variance V_ϕ; for small departures from the null hypothesis, the variance is approximately V'_ϕ or more crudely still, V_0.

In the following it should be remembered that we are investigating the judges, and not the underlying ranking. This was given *a priori*; each judge makes repeated attempts to reproduce it; we are concerned with

(*a*) estimating the degree to which a judge can reproduce the ranking, i.e. (assuming he operates according to the ϕ-model) estimating his ϕ;

(*b*) comparing two attempts at ranking;

(*c*) testing for differences between the attempts of several judges.

Estimation. Suppose a judge has produced l different rankings of the same n objects; it

is desired to estimate his ϕ. We shall do this by the method of maximum likelihood. We have from eqns. (12) and (16)

$$\log P\{(u)_1, (u)_2, \ldots, (u)_l\} = \log \phi \sum_{i=1}^{l} S(u)_i - l \log M_n(\phi),$$

whence the estimation equation is

$$\bar{S} = \phi \frac{d}{d\phi} \log M_n(\phi) \,|_{\phi=\hat{\phi}} = \mathscr{E}_{\hat{\phi}} = \mathscr{E}(S \mid \hat{\phi});$$

thus the model is fitted by its first moment. For small departures from $\mathscr{H}_0$ we have approximately

$$\bar{S} \simeq V_0 \log \hat{\phi} = V_0 \hat{\delta}.$$

We may give an approximate confidence interval for ϕ; since under $\mathscr{H}(\phi)$, $\bar{S}$ is approximately Normal with mean $\mathscr{E}_\phi$ and variance V'_ϕ/l, we have approximately

$$P\{\bar{S} - \lambda_\alpha \sqrt{(V'_\phi/l)} < \mathscr{E}_\phi < \bar{S} + \lambda_\alpha \sqrt{(V'_\phi/l)} \mid \phi\} \simeq 1 - 2\alpha,$$

where λ_α is the $100\alpha\,\%$ point of the unit Normal distribution; more crudely

$$P\{\bar{S} - \lambda_\alpha \sqrt{(\hat{V}'_\phi/l)} < V_0 \log \phi < \bar{S} + \lambda_\alpha \sqrt{(\hat{V}'_\phi/l)} \mid \phi\} \simeq 1 - 2\alpha,$$

where

$$\hat{V}'_\phi = V_0 - \frac{27}{25} \frac{\bar{S}^2}{n}. \tag{22}$$

Test for consistency between two judges. Suppose two judges have produced rankings (u_1) and (u_2), according to the ϕ-model, with respective parameters ϕ_1 and ϕ_2. Since $S(u)_1$ and $S(u)_2$ are approximately Normally distributed with means $V_0 \log \phi_1$ and $V_0 \log \phi_2$, and with variances V_0, we may test whether $\phi_1 = \phi_2$ by referring

$$(S(u)_1 - S(u)_2)/\sqrt{(2V_0)}$$

to Normal tables. An improved test is obtained by replacing V_0 by $\hat{V}'_\phi$ from (22); i.e. we refer

$$(S(u)_1 - S(u)_2) \Big/ \sqrt{\left\{2V_0 - \frac{27}{50n} (S(u)_1 + S(u)_2)^2\right\}}$$

to Normal tables.

This test is asymptotically equivalent to the likelihood ratio test.

Test for consistency between several judges. Suppose m judges have each produced l rankings with scores S_{ij} $(j = 1, 2, \ldots, l;\ i = 1, 2, \ldots, m)$. We associate with each judge a value ϕ_i of the parameter ϕ; then under the hypothesis $\phi_1 = \phi_2 = \ldots = \phi_m$, the 'between judges sum of squares'

$$l \sum_{i=1}^{m} (\bar{S}_{i.} - \bar{S}_{..})^2$$

(with the usual notation for means) is approximately distributed as $V_0 \chi^2_{m-1}$, and the 'within judges sum of squares'

$$\sum_{i=1}^{m} \sum_{j=1}^{l} (S_{ij} - \bar{S}_{i.})^2$$

is approximately distributed as $V_0 \chi^2_{m(l-1)}$. Thus we may test for differences between judges and for inconsistency within judges. More accurate tests would replace V_0 by the estimated $\hat{V}'_\phi$ as before.

The approximate powers of the above tests are immediately obtainable from standard Normal theory. Thus, for example, the power of the test to detect a difference between ϕ_1 and ϕ_2 when a two-tailed test is used is approximately

$$1-F(\lambda_\alpha-\mu)+F(-\lambda_\alpha-\mu),$$

where λ_α is as before the $100\alpha\,\%$ Normal point, and

$$\mu=\sqrt{(\tfrac{1}{2}V_0)}\,(\log\phi_1-\log\phi_2).$$

12. The conditional expectation of R

We now obtain an expression giving the conditional expectation of R, given S; this is the same for the ϕ-model as for the null case. Writing as in § 6 $M_0(x,y)$ for the null-hypothesis p.g.f. of R and S,

$$M_0(x,y)=\sum_{R,S}x^Ry^SP_0(R,S),$$

the p.g.f. for the ϕ-model is $M_0(x,y\phi)/M_0(1,\phi)$. Thus

$$\mathscr{E}_\phi(R)=\frac{\partial}{\partial x}\frac{M_0(x,\phi)}{M_0(1,\phi)}\bigg|_{x=1}. \tag{23}$$

However, we have for the null hypothesis

$$\frac{\partial}{\partial x}M_0(x,y)\big|_{x=1}=\sum_S y^SP_0(S)\,\mathscr{E}_0(R\,|\,S),$$

and so from (20) and (23)

$$\begin{aligned}\sum_S\phi^SP_0(S)\,\mathscr{E}_0(R\,|\,S)&=\mathscr{E}_\phi(R)\,M_0(1,\phi)\\&=(n!)^{-1}\sum_{2\leqslant k\leqslant n}(2k-n-1)\left\{k\,\frac{\phi^k+\phi^{-k}}{\phi^k-\phi^{-k}}-\frac{\phi+\phi^{-1}}{\phi-\phi^{-1}}\right\}\prod_{1\leqslant i\leqslant n}\frac{\phi^i-\phi^{-i}}{\phi-\phi^{-1}}\\&=(n!)^{-1}\sum_{2\leqslant k\leqslant n}(2k-n-1)\,\{(k-1)\,\phi^{k-1}+(k-3)\,\phi^{k-3}+\ldots+(1-k)\,\phi^{1-k}\}\\&\qquad\times\prod_{\substack{2\leqslant i\leqslant n\\ i\neq k}}\{\phi^{i-1}+\phi^{i-3}+\ldots+\phi^{1-i}\}.\end{aligned}$$

13. Conclusion

The ϕ-model exhibits several features of simplicity; however, difficulty has been encountered in attempting to evaluate such quantities as $P\{u_i=j\}$ and $P\{u_i=j,\,u_k=l\}$; knowledge of the latter quantity would enable the variance of R and covariance of R and S to be obtained; hence also the conditional variance of R, given S.

It is hoped in a further communication to present:

(*a*) an investigation of the result of § 12;

(*b*) a comparison of the present distributions of R and S with those obtained from other models, e.g. sampling from a bivariate Normal distribution;

(*c*) an extension of the method to cover those cases where the true ranking is not known *a priori*.

APPENDIX I

Demonstration that Kendall's t_k is not essentially a 'generalized matching coefficient'

To satisfy eqn. (6) with $C(u) = S(u)$ (the score for Kendall's t_k), we could take (for $n \leqslant 9$)

$$\alpha_{ij} = j \cdot 10^i$$

when $A(u)$ is simply the decimal number with digits u_i; e.g. $A(3, 5, 2, 4, 1) = 35{,}241$. We then define the function $F(x)$ suitably at the $n!$ different values of x (e.g. between 12,345 and 54,321) and have (6). However, no matter what $\{\alpha_{ij}\}$ are, we cannot satisfy the inverse relation (7) for all rankings (u) without having some of the values $F^{-1}(S) = g_S$ equal for different S. Thus with $n = 4$, consider the rankings with $u_1 = 1$:

(u)	$S(u)$	$F^{-1}(S(u))$		$A(u)$
(1, 2, 3, 4)	6	g_6	=	$\alpha_{11}+\alpha_{22}+\alpha_{33}+\alpha_{44}$
(1, 2, 4, 3)	4	g_4	=	$\alpha_{11}+\alpha_{22}+\alpha_{34}+\alpha_{43}$
(1, 3, 2, 4)	4	g_4	=	$\alpha_{11}+\alpha_{23}+\alpha_{32}+\alpha_{44}$
(1, 3, 4, 2)	2	g_2	=	$\alpha_{11}+\alpha_{23}+\alpha_{34}+\alpha_{42}$
(1, 4, 2, 3)	2	g_2	=	$\alpha_{11}+\alpha_{24}+\alpha_{32}+\alpha_{43}$
(1, 4, 3, 2)	0	g_0	=	$\alpha_{11}+\alpha_{24}+\alpha_{33}+\alpha_{42}$

Hence for any $\{\alpha_{ij}\}$ we must have

$$g_6 - 2g_4 + 2g_2 - g_0 = 0.$$

There are three similar equations obtainable from the rankings with $u_1 = 2, 3, 4$ in turn; together, they imply that $g_6 = g_{-6}$. *A fortiori*, (7) is impossible for $n > 4$.

APPENDIX II

To prove that Assumption B contains Assumption C we shall exhibit various pairs of inverse rankings. We shall use the 'cycle' notation; thus with $n = 6$, the cycle (1, 5, 6, 2) (or (5, 6, 2, 1) or (6, 2, 1, 5) or (2, 1, 5, 6)) will represent the ranking

$$u_1 = 5,\ u_5 = 6,\ u_6 = 2,\ u_2 = 1;\quad u_3 = 3,\ u_4 = 4;$$

i.e. the ranking (5, 1, 3, 4, 6, 2). The inverse ranking (2, 6, 3, 4, 1, 5) is represented by the cycle (2, 6, 5, 1), which is the reflexion of the first cycle. Thus any cycle whose reflexion is not equivalent to itself will specify a pair of inverse rankings. From the following cycles we now obtain as in § 5 the required relations between the λ_{ij}'s

Cycle	Relation		
To prove C (i):			
(1, 2, 3)	$\lambda_{12} = \lambda_{23}$		
(2, 3, 4)	$\lambda_{23} = \lambda_{34}$		
etc.		Hence	$\lambda_{i\,i+1} = \lambda_1 \quad (1 \leqslant i \leqslant n-1)$
(1, 2, 3, 4)	$\lambda_{12}\lambda_{13} = \lambda_{24}\lambda_{34}$	i.e.	$\lambda_{13} = \lambda_{24}$
(2, 3, 4, 5)	$\lambda_{23}\lambda_{24} = \lambda_{35}\lambda_{45}$	i.e.	$\lambda_{24} = \lambda_{35}$
etc.		Hence	$\lambda_{i\,i+2} = \lambda_2 \quad (1 \leqslant i \leqslant n-2)$
(1, 2, 3, 4, 5)	$\lambda_{12}\lambda_{13}\lambda_{14} = \lambda_{25}\lambda_{35}\lambda_{45}$	i.e.	$\lambda_{14} = \lambda_{25}$
etc.		Hence	$\lambda_{i\,i+3} = \lambda_3 \quad (1 \leqslant i \leqslant n-3)$
and so on.		Hence	$\lambda_{i\,i+k} = \lambda_k \quad (1 \leqslant i < i+k \leqslant n)$
To prove C (ii):			
(1, 3, 4, 2)	$\lambda_{13}\lambda_{23}\lambda_{24} = \lambda_{12}\lambda_{14}\lambda_{34}$	i.e.	$\lambda_1\lambda_3 = \lambda_2^2$
(1, 4, 5, 2)	$\lambda_2\lambda_4 = \lambda_2^2$		
(1, 5, 6, 2)	$\lambda_3\lambda_5 = \lambda_4^2$		
etc.		Hence	$\lambda_{k-1}\lambda_{k+1} = \lambda_k^2 \quad (2 \leqslant k \leqslant n-1)$

REFERENCES

BARTON, D. E. & DAVID, F. N. (1956). 'Spearman's Rho' and the matching problem. *Brit. J. Statist. Psychol.* **9**, 69–73.

BRADLEY, R. A. & TERRY, M. A. (1952). Rank analysis of incomplete block designs. I. *Biometrika*, **39**, 324–45.

DANIELS, H. E. (1944). The relation between measures of correlation in the universe of sample permutations. *Biometrika*, **33**, 129–35.

DANIELS, H. E. (1950). Rank correlation and population models. *J. R. Statist. Soc.* B, **12**, 171–81.

FELLER, W. (1945). The fundamental limit theorems in probability. *Bull. Amer. Math. Soc.* **51**, 800–32.

HADEN, H. G. (1947). A note on the distribution of the different orderings of n objects. *Proc. Camb. Phil. Soc.* **43**, 1–9.

HOEFFDING, W. (1948). A class of statistics with asymptotically Normal distribution. *Ann. Math. Statist.* **19**, 293–325.

HOTELLING, H. & PABST, M. R. (1936). Rank correlation and tests of significance involving no assumption of Normality. *Ann. Math. Statist.* **7**, 29–43.

KENDALL, M. G. (1950). Discussion on symposium on ranking methods. *J. R. Statist. Soc.* B, **12**, 189.

KENDALL, M. G. (1955). *Rank Correlation Methods.* London: Charles Griffin and Co. Ltd.

LEHMANN, E. L. (1953). The power of rank tests. *Ann. Math. Statist.* **24**, 23–43.

MORAN, P. A. P. (1950). Recent developments in ranking theory. *J. R. Statist. Soc.* B, **12**, 153–62.

MORAN, P. A. P. (1951). Partial and multiple rank correlation. *Biometrika*, **38**, 26–32.

SMITH, B. BABINGTON (1950). Discussion on Professor Ross's paper. *J. R. Statist. Soc.* B, **12**, 54.

STUART, A. (1954). The correlation between variate-values and ranks in samples from a continuous distribution. *Brit. J. Statist. Psychol.* **7**, 37–44.

THE GENERALIZATION OF PROBIT ANALYSIS TO THE CASE OF MULTIPLE RESPONSES

By J. AITCHISON and S. D. SILVEY
University of Glasgow

1. Introduction

The generalized probit analysis which we will discuss in this paper arose from a problem in entomology. We will suggest a solution to this particular problem and then consider a more general situation where this method of solution might be appropriate.

The particular problem is as follows. In the course of its lifetime the *Petrobius* Leach (Thysanura, Machilidae), which is a primitive wingless insect allied to the domestic silverfish, passes through various stages, technically referred to as 'instars'. A problem of biological interest is to estimate the mean time spent by such insects in each instar (stage). Since there are difficulties involved in keeping these insects alive in the laboratory (and consequently in keeping them under effectively continuous observation during their lifetime), they were sampled at various dates in the field, and those observed were classified according to instar. The actual set of data to be analysed is given in Table 1.* The insects started hatching on 30 April and times will be measured from this date. The problem is to estimate from these data the mean time spent in each stage.

Table 1. *Numbers of insects observed in various stages*

Date	Stage						Total
	1	2	3	4	5	6	
29 May	31	2	0	0	0	0	33
3 June	78	21	0	0	0	0	99
14 June	18	90	38	4	0	0	150
26 June	0	31	63	23	1	0	118
19 July	0	0	2	12	65	21	100

2. Basic mathematical model

For the purposes of a mathematical model of the situation outlined in the introduction we suppose that there are $s+1$ stages in the development of an insect, and that an insect observed is necessarily in one of these stages. The final $(s+1)$th stage is eventually reached by all insects and there is no question of estimating any unknown parameters for this stage. We also suppose that observations are made at m different times denoted by x_α $(\alpha = 1, 2, \ldots, m)$.

* For entomological details and sampling technique see Delany (1957). Stage 1 in Table 1 consists of hatching time together with three natural stages, since there seemed to be difficulty in distinguishing among these.

The time spent by an insect in the ith stage ($i = 1, 2, \ldots, s$) can be regarded as an observation on a non-negative random variable ξ_i, and our problem is to estimate the mean value λ_i of each ξ_i. Also the total time spent by an insect in stages $1, 2, \ldots, r$ can be regarded as an observation on a random variable $\eta_r = \sum_{i=1}^{r} \xi_i$. If $\mu_r = E(\eta_r)$ ($r = 1, 2, \ldots, s$), then clearly we can estimate λ_i ($i = 1, 2, \ldots, s$) by estimating μ_r ($r = 1, 2, \ldots, s$) and then taking differences of these estimates.

If, for each r, the distribution function G_r of η_r is continuous, so that

$$\Pr(\eta_r < x) = \Pr(\eta_r \leqslant x) = G_r(x),$$

then the probability $\pi_i(x)$ that at time x an insect chosen at random will be in the ith stage is given by

$$\begin{aligned} \pi_1(x) &= 1 - G_1(x), \\ \pi_i(x) &= G_{i-1}(x) - G_i(x) \quad (i = 2, 3, \ldots, s), \\ \pi_{s+1}(x) &= G_s(x). \end{aligned}$$

The expressions given for $\pi_1(x)$ and $\pi_{s+1}(x)$ are obvious. For $i = 2, 3, \ldots, s$, we have, in the usual notation,

$$\begin{aligned} \pi_i(x) &= \Pr(\eta_{i-1} < x, \eta_i > x) \\ &= \Pr(\eta_{i-1} < x) - \Pr(\eta_{i-1} < x, \eta_i \leqslant x) \\ &= \Pr(\eta_{i-1} < x) - \Pr(\eta_i < x), \end{aligned}$$

since the distribution functions of the η_i's are continuous and since $\eta_i < x$ implies $\eta_{i-1} < x$. These expressions for $\pi_i(x)$ ($i = 1, 2, \ldots, s+1$) enable us to express the likelihood of the observed data explicitly in terms of the distribution functions of the η_r's.

We now assume that, for $i = 1, 2, \ldots, s$, the distribution of the random variable ξ_i is approximately normal with standard deviation σ_i, small relative to λ_i. Then η_r also is approximately normally distributed and we may take $G_r(x) = \Phi(z_r) = \dfrac{1}{\sqrt{(2\pi)}} \displaystyle\int_{-\infty}^{z_r} e^{-\frac{1}{2}t^2}\, dt$, where $z_r = \dfrac{x - \mu_r}{\theta_r}$ and $\theta_r^2 = \operatorname{var}(\eta_r) = \operatorname{var}\left(\sum_{i=1}^{r} \xi_i\right)$. This assumption completes the basic model.

We note that the assumption that the standard deviation of ξ_i is small relative to its mean is a necessary adjunct of the assumption of approximate normality of ξ_i. It is an essential part of our model that ξ_i be a non-negative random variable. It can be non-negative and approximately normal only when its standard deviation is small relative to its mean, say, $3\sigma_i \leqslant \lambda_i$. It must be emphasized that in other applications this condition may not hold. Then it would be necessary to assume some other form for the distributions of the ξ_i's, which would make these effectively non-negative random variables.

3. Test of this model

In the above model, the only real assumption made is that of approximate normality of the ξ_i's. At this point a very easily applied test of this assumption is available. For subject to the assumption we have

$$\sum_{i=1}^{r} \pi_i(x) = 1 - \Phi(z_r) = \Phi(-z_r) = \Phi\left(\frac{\mu_r - x}{\theta_r}\right),$$

and so

$$\frac{\mu_r - x}{\theta_r} = \Phi^{-1}\left(\sum_{i=1}^{r} \pi_i(x)\right).$$

Now if $n_{\alpha i}$ denotes the number of insects observed in stage i at time x_α, N_α is the total number observed at time x_α, and $N_\alpha p_{\alpha i} = n_{\alpha i}$, then $p_{\alpha i}$ estimates $\pi_i(x_\alpha)$ and $\sum_{i=1}^{r} p_{\alpha i}$ estimates $\sum_{i=1}^{r} \pi_i(x_\alpha)$. Hence if $Y_{\alpha r} = \Phi^{-1}\left(\sum_{i=1}^{r} p_{\alpha i}\right)$, the equivalent normal deviate of $\sum_{i=1}^{r} p_{\alpha i}$, then, for fixed r, the points $(x_\alpha, Y_{\alpha r})$ $(\alpha = 1, 2, \ldots, m)$ will be well fitted by a straight line. By plotting these points and considering for each r their nearness to linearity we can decide whether or not the assumption of normality is justified and the model adequate.

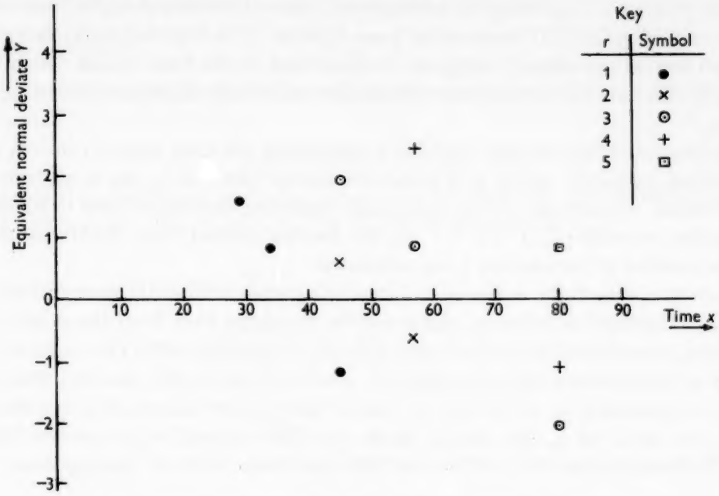


Fig. 1. Equivalent normal deviate of $\sum_{i=1}^{r} p_{\alpha i}$ plotted against time.

The diagram for the above set of data is given in Fig. 1. In considering this diagram it seems reasonable to give little weight to those points based on values of $\sum_{i=1}^{r} p_{\alpha i}$ near 0 or 1, since small variations in numbers observed can cause considerable movement of such points. It will be seen that for those values of r for which more than one 'reasonable' point is available, these points lie very near straight lines.

4. ESTIMATION

We are now in a position to estimate the unknown parameters μ_i $(i = 1, 2, \ldots, s)$ by the method of maximum likelihood. For now the likelihood of the observed set of data can be expressed explicitly in terms of the $2s$ unknown parameters μ_i and θ_i $(i = 1, 2, \ldots, s)$. While the $2s$ maximum-likelihood equations for their estimates are readily derived they are not simple and have to be solved numerically by an iterative procedure.

We remark at this point that the solution of these equations is tantamount to the fitting of straight lines to each of the sets of points $(x_\alpha, Y_{\alpha r})$ $(\alpha = 1, 2, \ldots, m)$. Initial approximations to the maximum-likelihood estimates can be obtained by fitting straight lines roughly to these sets of points. For given r a straight line fitted thus to the points $(x_\alpha, Y_{\alpha r})$

($\alpha = 1, 2, \ldots, m$) will cross the x-axis near the maximum-likelihood estimate of μ_r, while the gradient will approximate to the maximum-likelihood estimate of $-\theta_r^{-1}$.

However, we omit further details of this case since these are similar tc those given in § 6, where the procedure of obtaining estimates using a modified model is demonstrated for the data of § 1.

5. Modifications of the basic model

It will be seen from the discussion in § 4 that the given data are really insufficient for good estimation of $\mu_1, \mu_2, \ldots, \mu_5$, using the basic model, since of the five straight lines to be fitted only three are being fitted to 'reasonable' sets of points. This, together with other considerations based less on expediency, suggests modifications to the basic model which involve a reduction of the number of unknown parameters and which might prove useful in similar situations.

The variance σ_i^2 of the random variable ξ_i describing the time spent in the ith stage has been regarded, up to this point, as a parameter independent of λ_i, the mean time spent in stage i. Further, no assumption has been made regarding the dependence or independence of the random variables ξ_i ($i = 1, 2, \ldots, s$). By making assumptions of this nature we can reduce the number of parameters to be estimated.

There are two aims to be achieved in introducing such additional assumptions: first, to make them as natural as possible, and secondly, to ensure that from the point of view of computation, the estimation problem they give rise to is practicable. The 'natural' assumptions make σ_i^2 a simple function of the mean λ_i, whereas in order that the computation should not be too complicated, we would like to assume that θ_r^2, the variance of η_r, is a simple function of μ_r, the mean of η_r. Sometimes these aims are compatible; sometimes they are in conflict. To illustrate this we consider the following three 'natural' assumptions

$$\text{(i)} \quad \sigma_i^2 = \sigma^2,$$

$$\text{(ii)} \quad \sigma_i^2 = \sigma^2 \lambda_i,$$

$$\text{(iii)} \quad \sigma_i^2 = \sigma^2 \lambda_i^2,$$

where in each case σ^2 is a constant. If now we further assume independence of the ξ_i's, i.e. if we assume that the time spent by an insect in any one stage is independent of the time spent by it in any other stage, then, in terms of θ_r and μ_r, these assumptions become

$$\text{(i)}' \quad \theta_r^2 = r\sigma^2,$$

$$\text{(ii)}' \quad \theta_r^2 = \mu_r \sigma^2,$$

$$\text{(iii)}' \quad \theta_r^2 = \left[\mu_1^2 + \sum_{i=2}^{r} (\mu_i - \mu_{i-1})^2\right] \sigma^2.$$

From both points of view the assumptions (i)' and (ii)' are 'good'. However, while (iii)' is a very natural assumption it leads to a very unwieldy estimation problem. We would like to be able to replace (iii)' by $\theta_r^2 = \mu_r^2 \sigma^2$. But this is an artificial assumption which would be consistent with the assumption (iii) only if we assume very high correlation between any two of the ξ_i's.

The assumption which seemed to the authors most reasonable for the given set of data was the assumption (ii) in conjunction with independence of the ξ_i's, i.e. essentially the

assumption (ii)′. An initial test of whether such an assumption is valid is obtained by reference to Fig. 1. For if the assumption is valid then a straight line fitted to the set of points $(x_\alpha, Y_{\alpha r})$, r fixed ($\alpha = 1, 2, \ldots, m$), should approximate to the line with equation $Y = \dfrac{\mu_r - x}{\sigma\sqrt{\mu_r}}$. From the point where this fitted line crosses the x-axis and from its gradient, an estimate of σ may be obtained, without difficulty. If, as r varies, the different estimates of σ thus obtained are in fairly close agreement and show no definite trend as r increases, then the additional assumption (ii)′ may be taken to be justified. This was found to be the case for the given set of data, if again not too much weight is attached to points based on values of $\sum_{i=1}^{r} p_{\alpha i}$ near 0 or near 1.

Usually it will be possible to test an additional assumption of this nature in a similar fashion.

6. APPLICATION TO GIVEN DATA

Summarizing the model used to fit the given data we have: the probability $\pi_i(x_\alpha)$ that at time x_α an insect chosen at random will be in the ith stage is given by

$$\pi_1(x_\alpha) = 1 - \Phi\left(\frac{x_\alpha - \mu_1}{\sigma\sqrt{\mu_1}}\right),$$

$$\pi_i(x_\alpha) = \Phi\left(\frac{x_\alpha - \mu_{i-1}}{\sigma\sqrt{\mu_{i-1}}}\right) - \Phi\left(\frac{x_\alpha - \mu_i}{\sigma\sqrt{\mu_i}}\right) \quad (i = 2, 3, \ldots, s),$$

$$\pi_{s+1}(x_\alpha) = \Phi\left(\frac{x_\alpha - \mu_s}{\sigma\sqrt{\mu_s}}\right),$$

where $\Phi(z) = \dfrac{1}{\sqrt{(2\pi)}}\displaystyle\int_{-\infty}^{z} e^{-\frac{1}{2}t^2}\,dt$. Also the likelihood of the observed data is given by

$$\log L = k + \sum_{\alpha=1}^{m}\sum_{i=1}^{s+1} n_{\alpha i} \log \pi_i(x_\alpha),$$

where k is a constant and $n_{\alpha i}$ is the number of insects observed in stage i at time x_α.

We now have

$$\frac{\partial \pi_i(x_\alpha)}{\partial \mu_i} = -\frac{\partial \pi_{i+1}(x_\alpha)}{\partial \mu_i} = \frac{1}{2\mu_i}\frac{x_\alpha + \mu_i}{\sigma\sqrt{\mu_i}} Z_{\alpha i} \quad (i = 1, 2, \ldots, s),$$

where $$Z = \frac{1}{\sqrt{(2\pi)}} e^{-\frac{1}{2}z^2} \quad \text{and} \quad z_{\alpha i} = \frac{x_\alpha - \mu_i}{\sigma\sqrt{\mu_i}}.$$

Also $$\frac{\partial \log L}{\partial \mu_i} = \sum_{\alpha=1}^{m}\left(\frac{n_{\alpha i}}{\pi_i(x_\alpha)} - \frac{n_{\alpha i+1}}{\pi_{i+1}(x_\alpha)}\right)\frac{\partial \pi_i(x_\alpha)}{\partial \mu_i} \quad (i = 1, 2, \ldots, s).$$

Further $$\frac{\partial \pi_i(x_\alpha)}{\partial \sigma} = \frac{1}{\sigma}[z_{\alpha i} Z_{\alpha i} - z_{\alpha i-1} Z_{\alpha i-1}] \quad (i = 2, 3, \ldots, s)$$

and $$\frac{\partial \log L}{\partial \sigma} = \frac{1}{\sigma}\sum_{\alpha=1}^{m}\sum_{i=1}^{s}\left[\frac{n_{\alpha i}}{\pi_i(x_\alpha)} - \frac{n_{\alpha i+1}}{\pi_{i+1}(x_\alpha)}\right] z_{\alpha i} Z_{\alpha i}.$$

The problem of estimating the μ_i's is thus reduced to the problem of solving a set of $s+1$ equations each of which involves only functions which are either tabulated or easily calculated. (The only tabulated functions which it was found necessary to use were tables of the normal ordinate and integral and the computational procedure follows naturally from the form in which the formulae are given.)

These equations were solved in the present instance by an iterative procedure which is essentially that used in probit analysis (see, for example, Finney, 1947). This involves (i) obtaining initial approximations to the roots of the equations, (ii) evaluating **W**, the information matrix with the unknown parameters involved replaced by these initial approximations, (iii) evaluating $\mathbf{W}^{-1}$ and (iv) obtaining corrections to the initial approximation by multiplying by $\mathbf{W}^{-1}$ the vector of partial derivatives of $\log L$ with respect to the parameters, calculated with these parameters replaced by the initial approximations.

This method is a modified form of Newton's method of solving numerically a set of simultaneous equations and has the advantage over the latter method that it yields finally, without additional computation, an estimate of the variance matrix of the estimators. Also the matrix is more easily computed than that used in Newton's method.

We now discuss the points (i) and (ii) for our particular case.

(i) We emphasize again that the solution of the maximum-likelihood equations is tantamount to the fitting of straight lines to the sets of points $(x_\alpha, Y_{\alpha r})$, r fixed ($\alpha = 1, 2, \ldots, m$). Hence initial approximations to the roots of these equations can be obtained by fitting straight lines roughly to these sets of points. In order to reduce the number of iterations required in the process of solution it seems advisable to take considerable care with the fitting of these lines, and in so far as possible to attach weights to different points on the basis previously discussed. Moreover, scrutiny of this diagram can yield considerable information about how well we can expect different parameters to be estimated. It is obvious, for instance, that in the present case we can expect μ_1, μ_2 and μ_3 to be well estimated and the remaining μ's not so well estimated. Hence before any heavy computations are carried out the experimenter can decide whether the experiment he has performed is likely to yield as much information as he desires or whether it might be necessary, for example, to repeat the experiment with observations made at closer intervals.

Fig. 2 shows the initial lines fitted in the present case. The lines resulting from the final solution of the maximum-likelihood equations are virtually indistinguishable from these. The corresponding initial and final estimates of the parameters are given in Table 2. Two iterations were required to produce the final solution of the maximum-likelihood equations. It will be seen that these produce little change in the initial approximations to the estimates of the μ_i's, the parameters in which we are primarily interested. So graphical estimation of these parameters may be quite efficient.

Table 2

Parameter	μ_1	μ_2	μ_3	μ_4	μ_5	σ
Initial approximation	38·0	50·7	62·7	72·0	88·0	1·00
Maximum-likelihood estimate	38·3	50·9	62·5	71·9	87·9	1·01

(ii) For the evaluation of **W** we have

$$E\left(-\frac{\partial^2 \log L}{\partial \mu_i^2}\right) = \sum_{\alpha=1}^{m} N_\alpha \left[\frac{1}{\pi_i(x_\alpha)} + \frac{1}{\pi_{i+1}(x_\alpha)}\right]\left[\frac{\partial \pi_i(x_\alpha)}{\partial \mu_i}\right]^2 \quad (i = 1, 2, \ldots, s),$$

$$E\left(-\frac{\partial^2 \log L}{\partial \mu_i \partial \mu_{i+1}}\right) = -\sum_{\alpha=1}^{m} \frac{N_\alpha}{\pi_i(x_\alpha)} \frac{\partial \pi_i(x_\alpha)}{\partial \mu_i} \frac{\partial \pi_{i+1}(x_\alpha)}{\partial \mu_{i+1}} \quad (i = 1, 2, \ldots, s-1),$$

and

$$E\left(-\frac{\partial^2 \log L}{\partial \mu_i \partial \mu_j}\right) = 0 \quad \text{if} \quad |i-j| \geqslant 2.$$

Also $$E\left(-\frac{\partial^2 \log L}{\partial \sigma^2}\right) = \frac{1}{\sigma}\sum_{\alpha=1}^{m}\sum_{i=1}^{s}\left[\frac{N_\alpha}{\pi_i(x_\alpha)}\frac{\partial \pi_i(x_\alpha)}{\partial \sigma} - \frac{N_\alpha}{\pi_{i+1}(x_\alpha)}\frac{\partial \pi_{i+1}(x_\alpha)}{\partial \sigma}\right] z_{\alpha i} Z_{\alpha i},$$

and $$E\left(-\frac{\partial^2 \log L}{\partial \mu_i \partial \sigma}\right) = \sum_{\alpha=1}^{m}\left[\frac{N_\alpha}{\pi_i(x_\alpha)}\frac{\partial \pi_i(x_\alpha)}{\partial \sigma} - \frac{N_\alpha}{\pi_{i+1}(x_\alpha)}\frac{\partial \pi_{i+1}(x_\alpha)}{\partial \sigma}\right]\frac{\partial \pi_i(x_\alpha)}{\partial \mu_i} \quad (i = 1, 2, \ldots, s).$$

In the computation of these expectations it is necessary to calculate the $1/[\pi_i(x_\alpha)]$'s. Sometimes $\pi_i(x_\alpha)$ will be zero to the accuracy of the calculations. The terms in which such

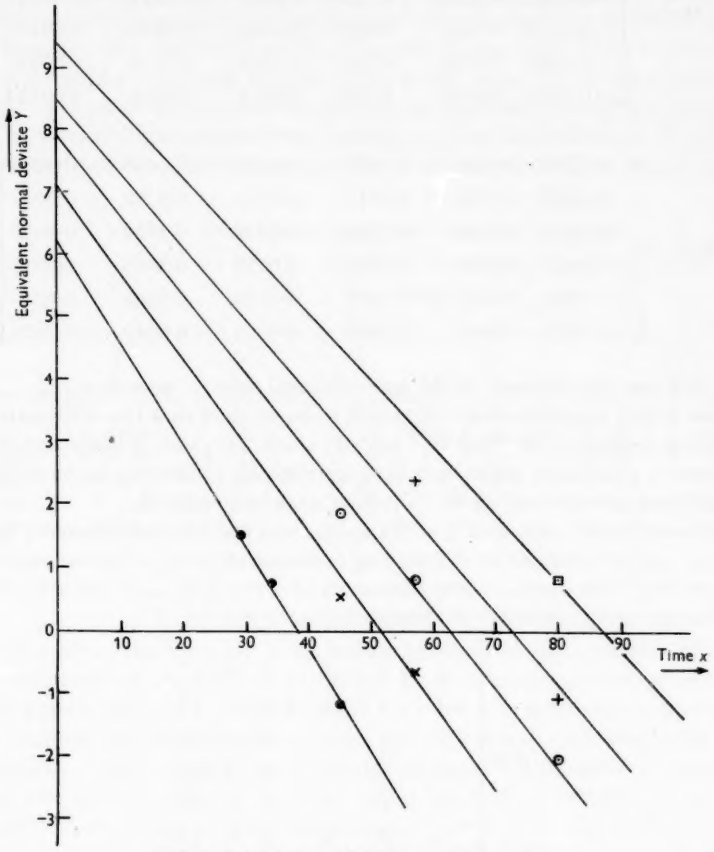


Fig. 2. Initial lines fitted to points $(x_\alpha, Y_{\alpha r})$.

$1/[\pi_i(x_\alpha)]$'s are involved will invariably be negligible and as a working rule we may take $1/[\pi_i(x_\alpha)] = 0$ whenever $\pi_i(x_\alpha) = 0$.

The information matrix takes a simple form in this case and advantage may be taken of the empty cells in it in the inversion process. The authors inverted it by successive bordering (Frazer, Duncan & Collar, 1947, p. 112).

The matrix $\mathbf{W}^{-1}$ is an estimate of the variance matrix $\mathbf{V}$ of the maximum-likelihood estimators of the unknown parameters and $\mathbf{W}^{-1}$ may be used without alteration in successive iterations to obtain corrections to successive approximations to maximum-likelihood

estimates. However, a better estimate of $\mathbf{V}$ may be obtained finally by computing $\mathbf{W}^{-1}$, the information matrix with unknown parameters replaced by their maximum-likelihood estimates, and using $\mathbf{W}_1^{-1}$ as an estimate of $\mathbf{V}$. For the given set of data we have

$$\mathbf{W}^{-1} = \begin{bmatrix} 0{\cdot}3080 & 0{\cdot}0419 & -0{\cdot}0065 & 0{\cdot}0000 & -0{\cdot}0530 & -0{\cdot}0050 \\ 0{\cdot}0419 & 0{\cdot}3890 & 0{\cdot}1156 & 0{\cdot}0333 & 0{\cdot}0161 & 0{\cdot}0006 \\ -0{\cdot}0065 & 0{\cdot}1156 & 0{\cdot}7835 & 0{\cdot}2238 & 0{\cdot}1521 & 0{\cdot}0084 \\ 0{\cdot}0000 & 0{\cdot}0333 & 0{\cdot}2238 & 0{\cdot}9988 & 0{\cdot}2942 & 0{\cdot}0016 \\ -0{\cdot}0530 & 0{\cdot}0161 & 0{\cdot}1521 & 0{\cdot}2942 & 2{\cdot}2144 & 0{\cdot}0239 \\ -0{\cdot}0050 & 0{\cdot}0006 & 0{\cdot}0084 & 0{\cdot}0016 & 0{\cdot}0239 & 0{\cdot}0022 \end{bmatrix},$$

while

$$\mathbf{W}_1^{-1} = \begin{bmatrix} 0{\cdot}3090 & 0{\cdot}0423 & -0{\cdot}0078 & -0{\cdot}0013 & -0{\cdot}0608 & -0{\cdot}0058 \\ 0{\cdot}0423 & 0{\cdot}3742 & 0{\cdot}1218 & 0{\cdot}0361 & 0{\cdot}0229 & 0{\cdot}0012 \\ -0{\cdot}0078 & 0{\cdot}1218 & 0{\cdot}7793 & 0{\cdot}2308 & 0{\cdot}1752 & 0{\cdot}0088 \\ -0{\cdot}0013 & 0{\cdot}0361 & 0{\cdot}2308 & 1{\cdot}0123 & 0{\cdot}3056 & 0{\cdot}0022 \\ -0{\cdot}0608 & 0{\cdot}0229 & 0{\cdot}1572 & 0{\cdot}3056 & 2{\cdot}2356 & 0{\cdot}0249 \\ -0{\cdot}0058 & 0{\cdot}0012 & 0{\cdot}0088 & 0{\cdot}0022 & 0{\cdot}0249 & 0{\cdot}0023 \end{bmatrix},$$

where in each case the elements of the main diagonal refer in order to $\mu_1, \mu_2, \ldots, \mu_5$ and σ. In this case initial approximations happened to be so good that the differences between corresponding elements of $\mathbf{W}^{-1}$ and $\mathbf{W}_1^{-1}$ are very small. However, if initial approximations had not been so good there might have been considerable differences in the corresponding elements of these matrices and so $\mathbf{W}_1^{-1}$ should always be computed.

As mentioned above, estimates $\hat{\lambda}_i$ of the mean times spent in the different stages, i.e. of λ_i $(i = 1, 2, \ldots, 5)$ are obtained by differencing estimates of the μ_i's. Also an estimate of the variance matrix of the corresponding estimators of λ_i $(i = 1, 2, \ldots, 5)$ and σ is obtained by forming the matrix $\mathbf{V}_{\lambda\,\sigma} = \mathbf{A}\mathbf{W}_1^{-1}\mathbf{A}'$, where

$$\mathbf{A} = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ -1 & 1 & 0 & 0 & 0 & 0 \\ 0 & -1 & 1 & 0 & 0 & 0 \\ 0 & 0 & -1 & 1 & 0 & 0 \\ 0 & 0 & 0 & -1 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix}.$$

and $\mathbf{A}'$ is the transpose of $\mathbf{A}$.

In the present instance we find that

$$\begin{aligned} \hat{\lambda}_1 &= 38{\cdot}3 \\ \hat{\lambda}_2 &= 12{\cdot}6 \\ \hat{\lambda}_3 &= 11{\cdot}6 \\ \hat{\lambda}_4 &= 9{\cdot}4 \\ \hat{\lambda}_5 &= 16{\cdot}0 \\ \hat{\sigma} &= 1{\cdot}01 \end{aligned} \qquad \mathbf{V}_{\lambda,\sigma} = \begin{bmatrix} 0{\cdot}309 & -0{\cdot}267 & -0{\cdot}050 & 0{\cdot}006 & -0{\cdot}060 & -0{\cdot}006 \\ -0{\cdot}267 & 0{\cdot}572 & -0{\cdot}202 & -0{\cdot}092 & 0{\cdot}046 & 0{\cdot}007 \\ -0{\cdot}050 & -0{\cdot}202 & 0{\cdot}910 & -0{\cdot}463 & -0{\cdot}060 & 0{\cdot}008 \\ 0{\cdot}006 & -0{\cdot}092 & -0{\cdot}463 & 1{\cdot}330 & -0{\cdot}633 & -0{\cdot}007 \\ -0{\cdot}060 & 0{\cdot}046 & -0{\cdot}060 & -0{\cdot}633 & 2{\cdot}637 & 0{\cdot}023 \\ -0{\cdot}006 & 0{\cdot}007 & 0{\cdot}008 & -0{\cdot}007 & 0{\cdot}023 & 0{\cdot}003 \end{bmatrix}.$$

It will be seen that the estimates of the λ's and σ are such that they do not invalidate our assumption that the standard deviation of each ξ_i is small relative to its mean.

If we assume that samples are large it is now possible to obtain, in the usual way, confidence intervals for the mean times spent in the different stages. Also, for example, if we had data on different batches of insects we could clearly test whether the times spent in the same stage by different batches were significantly different from one another.

We conclude this paragraph by comparing in Table 3 the expected number of insects in each stage at each time (when unknown parameters are replaced by the above estimates), with the corresponding observed number. If entries in this table where expectations are less than 5 are pooled with neighbouring entries at the same date in the obvious way (so that we have a total of twelve observed and expected numbers to compare) then a χ^2-test shows no significant differences between observed and expected numbers. Since a total of six parameters has been estimated the number of degrees of freedom associated with χ^2 is $12-6=6$.

Table 3. *Observed and (expected) numbers of insects*

Date	Stage					
	1	2	3	4	5	6
29 May	31 (30·8)	2 (2·2)	0 (0)	0 (0)	0 (0)	0 (0)
3 June	78 (74·8)	21 (23·3)	0 (0·9)	0 (0)	0 (0)	0 (0)
14 June	18 (21·3)	90 (97·8)	38 (28·8)	4 (2·0)	0 (0·1)	0 (0)
26 June	0 (0·2)	31 (23·2)	63 (65·6)	23 (24·1)	1 (4·8)	0 (0·1)
19 July	0 (0)	0 (0)	2 (1·4)	12 (15·7)	65 (62·6)	21 (20·3)

7. GENERALIZATION OF PROBIT ANALYSIS

The situation where the methods of quantal probit analysis are applicable is, in the customary terminology of this analysis, as follows. Random samples of *subjects* are subjected to various doses of a *stimulus* to each of which a subject may or may not respond. Any dose results in a dichotomy of the subjects to which it is applied into those responding and those not responding. A dose is said to be *effective* for a subject if it produces a response in that subject and the minimum effective dose for a subject is called its *tolerance*. One of the main objects of this type of analysis is to estimate mean tolerance (Finney, 1947).

Clearly a situation might arise where in place of a simple dichotomy, subjects are divided into more than two classes by any dose of the stimulus. Accordingly we envisage an experiment where random samples of subjects are subjected to m doses x_α ($\alpha = 1, 2, \ldots, m$) of a stimulus and as a result of the application of the dose x_α each subject is placed in one of $s+1$ classes. A straightforward illustration of such an experiment is given by Tattersfield, Gimingham & Morris (1925) who classified insects subjected to a poison as *unaffected*, *slightly affected*, *moribund* or *dead*. The particular problem discussed above is another illustration if, in this case, time is regarded as the stimulus.

We now ask what conditions must be satisfied in this general experiment in order that the method of analysis used in our particular case should be applicable.

Study of § 2 shows that the crucial conditions are

(i) the classes must be *ordered*, mutually exclusive and exhaustive,

(ii) the reactions of a subject to increasing doses must be systematic in the sense that if dose x places a subject in the ith class then a dose greater than x is required to place this subject in the jth class whenever j is greater than i.

When these conditions are satisfied then the model of § 2 can be used to describe the experimental situation in the following manner. If for a subject chosen at random the tolerance of this subject for the ith class is y_i ($i = 1, 2, \ldots, s$) (i.e. y_i is the minimum dose required to place this subject in the $(i+1)$th class) and if every subject is placed in the first class by dose $y_0 = 0$, then $y_i - y_{i-1}$ can be regarded as an observation on a *non-negative* random variable ξ_i ($i = 1, 2, \ldots, s$), which describes the *marginal tolerances* of the subjects for the ith class (i.e. the differences in tolerances for the ith and $(i-1)$th classes). The mean value λ_i of ξ_i is then the mean marginal tolerance for the ith class. Further, the random variable $\eta_r = \sum_{i=1}^{r} \xi_i$ describes the tolerances of subjects for the rth class, and $\mu_r = E(\eta_r)$ is the mean tolerance for the rth class. Methods similar to that given subsequently to § 2 may then be applied to estimate the μ_i's and/or λ_i's, whichever happen to be of biological interest. We emphasize again that the assumption of normality of the ξ_i's which we have used in our illustration is possible only if it can be assumed that the standard deviation of each ξ_i is small compared with its mean.

We remark finally that if $s = 1$ then the present analysis becomes an ordinary probit analysis and it is in this sense that we have generalized probit analysis.

REFERENCES

DELANY, M. J. (1957). The life history and ecology of the genus *Petrobius* Leach (Thysanura, Machilidae) (in the Press).

FINNEY, D. J. (1947). *Probit Analysis.* Cambridge University Press.

FRAZER, R. A., DUNCAN, W. J. & COLLAR, A. R. (1947). *Elementary Matrices and Some Applications to Dynamics and Differential Equations.* Cambridge University Press.

TATTERSFIELD, F., GIMINGHAM, C. T. & MORRIS, H. M. (1925). *Ann. Appl. Biol.* **12**, 61.

EXPERIMENTING WITH ORGANISMS AS BLOCKS

By S. C. PEARCE
East Malling Research Station

An experimenter may usually assume that his treatments will have little effect outside the plots† to which each is applied, but sometimes this is not justified. Thus, if an animal is injected at several places with different strains of inoculum it is conceivable that a strain that produces a large lesion locally will aggravate lesions elsewhere. Again, a tree may have each branch pruned according to a different method, and it is possible that a method which encourages cropping on the branch to which it is applied will encourage cropping elsewhere on the tree, though to a smaller degree. Nor are remote effects always of the same sign as the local ones. Thus, if the blossoms of a plant are grouped and each group is pollinated with a different variety, it may well be that a kind of pollen which induces high fruit-set locally may thereby inhibit fruiting elsewhere.

The existence and importance of such remote effects is often problematical, and may require to be investigated. Experience with the method to be described has shown that there are sometimes remote effects of considerable biological interest. On the other hand, it may appear that none exist, and simpler methods can then be used with confidence in the future.

Notation

The local effect of a treatment on the plots to which it is applied will be represented by the parameter γ_j, where j is the number of the treatment. Its remote effect on the other plots of the same block will be written δ_j. It will be assumed that it has no effects in other blocks, which will consist of different organisms. It will also be assumed that the remote effect is the same on all other plots of the block; this is a biological problem, but one that is fairly easy to decide, because the answer can be found in the physiological symmetry or asymmetry of the organism.

There is also the possibility of local and remote effects interacting. Thus, although it has been assumed that the remote effect of treatment j on a given plot does not depend upon the plot itself, it might depend upon the treatment that is being applied locally. The parameter ϕ_{kj}, will be taken as the additional effect of a plot receiving treatment k when treatment j is being applied elsewhere in the block.

There are instances in which this concept of local and remote effects may not appeal to the biologist. Thus, in the studies of the effects of different kinds of tuberculin described by Paterson & Leech (1954), the concept was rather of a systemic effect over the whole guinea-pig irrespective of the site of injection, and of a local effect at the site. Since, however, the systemic effect equals δ_j in the notation of this paper, and the local effect as now defined equals $(\gamma_j - \delta_j)$, little modification of the working is required.

The general parameter will be written as α, and the parameter for block h as β_h.

† The word 'plot' is used because it is the usual statistical term for the unit to which experimental treatments are applied. However, for trials of the kind to be considered these units will be very dissimilar from 'plots' in an agricultural sense.

The following equations of constraint will be used:

$$0 = \sum_h \beta_h = \sum_j \gamma_j = \sum_j \delta_j = \sum_k \phi_{kj} = \sum \phi_{kj}.$$

The choice of experimental designs

If an orthogonal design, e.g. randomized blocks, is used it is apparent that there is no means of separating the parameters of the remote effects, the evaluation of which is important. Thus, within a block all remote effects occur equally often and will be lost by virtue of the equation of constraint involving them. Nor will the parameters γ_j and δ_j be separable in the expressions for treatment totals; thus, if injection of a branch of a tree with a nutrient solution increases the crop as compared with the other branches by 30 lb., it is not possible to say whether there has been a local effect of $+30$ and no remote effect, or a local effect of $+50$ and a remote effect of $+20$, or what. For all the figures prove there might have been a local effect of -30 and a remote effect of -60.

If, however, balanced incomplete block designs are used, it is apparent that treatment comparisons within blocks will provide estimates of $(\gamma_j - \delta_j)$ for all j, as before; whereas comparisons between blocks will give estimates of $[\gamma_j + (k-1)\,\delta_j]$, where k is the number of plots to a block. Provided that the inter-block differences are determined with reasonable accuracy, it is thus possible to estimate γ_j and δ_j separately to a useful degree of precision. The design proposed is, in effect, one of b kinds of block, with v treatments, k plots to a block, and n of each kind of block so that there are nb blocks in all. There is no necessity for n to exceed one, but in practice it is usually desirable if the inter-block comparisons are to be of use.

Where it is required to take account of the interactions between local and remote effects, choice must be restricted to a narrower range of designs. Thus if the normal equations are to be manageable the design must be unreduced,† i.e. all combinations of the v treatments taken k at a time must occur equally often in blocks, so b must equal $\binom{v}{k}$. Also, if there are to be any degrees of freedom for error and if $k = v-1$, n must be at least two. Also, it will appear that k must be at least three, which means that v must be at least four.

Analysis of data

(i) *With interactions ignored*

This is not difficult. Let the grand total of all data be G, let the total of data from treatment j be T_j, and let the total from all blocks that do not contain treatment j be U_j.

† It has been pointed out to the writer by Mr G. H. Freeman that this restriction to unreduced designs is sufficient to ensure that no normal equation shall contain more than two ϕ parameters (i.e. ϕ_{ij} and ϕ_{jl}), but it is not necessary. The necessary restriction is this: Considering only blocks that contain treatment i but not treatment j, and omitting from them all plots to receive treatment i, then, for all i and j, the remaining design must be in balanced incomplete blocks. Let the parameters of the remaining design be given the conventional symbols, b, k, r, v and λ, and let the corresponding capitals refer to the original design, then

$$B = b(v+1)\,(v+2)/(v-k+1)\,(k+1), \quad K = k+1, \quad R = b(v+1)/(v-k+1),$$
$$V = v+2 \quad \text{and} \quad \Lambda = bk/(v-k+1).$$

The simplest integral solution of these equations, apart from those arising from unreduced designs, appears to be $b = 18$, $k = 4$, $r = 8$, $v = 9$, $\lambda = 3$, $B = 66$, $K = 5$, $R = 30$, $V = 11$ and $\Lambda = 12$. It is not known if such designs exist, but if they do they are too large to be of practical importance, and their possible existence does not justify complicating the algebra. Nevertheless, Mr Freeman's comment is of some theoretical interest.

Estimating treatment effects within blocks gives $(\gamma_j - \delta_j)$ equal to

$$\frac{(v-1)}{nbk(k-1)}(kT_j + U_j - G).$$

Estimating them between blocks gives $[\gamma_j + (k-1)\delta_j]$ as

$$\frac{(v-1)}{nbk(v-k)}[(v-k)G - vU_j].$$

Hence, estimate of γ_j equals

$$\frac{(v-1)}{nbk(v-k)}[(v-k)T_j - U_j],$$

and estimate of δ_j equals

$$\frac{(v-1)}{nbk(k-1)(v-k)}[(v-k)G - (v-1)U_j - (v-k)T_j].$$

The standard errors of these estimates are likewise readily obtained. If M be the error mean square of the analysis between blocks and M' that of the analysis within, and writing

$$M^0 = M + \frac{(k-1)(v-k)}{v}M',$$

and

$$M^+ = M + \frac{(v-k)}{v(k-1)}M',$$

then

$$[\text{standard error of estimate of } \gamma_j]^2 = \frac{(v-1)^2}{nbk^2(v-k)}M^0,$$

$$[\text{standard error of estimate of } (\gamma_i - \gamma_j)]^2 = \frac{2v(v-1)}{nbk^2(v-k)}M^0.$$

The corresponding expressions for the standard errors of δ_j and $(\delta_i - \delta_j)$ are the same except for the replacement of M^0 by M^+. If k equals two, each treatment gives rise to the parameter for its remote effect as often as to the one for its local effect, and M^0 equals M^+. As k increases from two, M^+ becomes progressively smaller than M^0 and the remote effects are determined with ever greater accuracy as compared with the local effects. This is just as well, because the remote effects are likely to be the smaller and so accordingly are differences between them.

If the treatments had been applied to whole organisms, estimates would have been obtained of the quantities $[\alpha + \gamma_j + (k-1)\delta_j]$. These values can also be obtained by adding their components; if this is done, the (standard error)^2 of the difference between two such estimates is

$$\frac{2v(v-1)M}{nb(v-k)}.$$

It should be noted that just as the parameter γ_j has been adjusted to $(\gamma_j - \delta_j)$ in the analysis within blocks, so also has β_h been adjusted to $\beta_h^* = \beta_h - \Sigma\delta_j$, summation taking place over j for treatments lacking in block h.

(ii) *With interactions allowed for*

Where interactions are considered the equations already given still apply, because the expressions for T_j, U_j and G will contain no parameters of the type ϕ_{jl}. It follows that if

ϕ_{jt} is to be evaluated use will have to be made of further quantities; the following will be found convenient:

D_{jt}, the total for treatment j in blocks from which treatment t is absent.

E_{jt} $(= E_{tj})$, the total for blocks containing neither treatment j nor t.

The total for blocks not containing treatment t; this is of course U_t.

Thus, D_{jt}, E_{jt} and U_t are respectively analogues of T_j, U_j and G. It will be convenient to write

$$F_{jt} = kD_{jt} + E_{jt} - U_t \quad \text{and} \quad V_j = kT_j + U_j - G.$$

The parametric expression for F_{jt} is

$$n\binom{v-3}{k-2}\left[\frac{(v-k-1)}{(v-3)}\phi_{tj} - \frac{(kv-2k-v+1)}{(v-3)}\phi_{jt} + (v-1)(\gamma_j - \delta_j) + (\gamma_t - \delta_t)\right].$$

Since, as has been shown, the corresponding expression for V_j is

$$\frac{nbk(k-1)}{(v-1)}(\gamma_j - \delta_j),$$

it follows that

$$(v-1)(v-2)(v-k)[(k-1)V_j + V_t] - k(v-k)[V_j + V_t] \\ - v(v-1)(v-2)[(k-1)F_{jt} + F_{tj}] + kv(v-2)[F_{jt} + F_{tj}]$$

is an estimate of $$nk(k-2)(v-k)v(v-1)\binom{v-2}{k-2}\phi_{jt}.$$

This expression can be set out in other ways, but it does not appear that any of the alternatives is more convenient from the point of view of practical computation, except when $k = v-1$, when it is easier to take

$$(v-1)V_j + V_t - v(v-2)F_{jt}$$

as an estimate of $nv(v-1)(v-2)\phi_{jt}$. It may be noted that in another special case, i.e. where $k = 2$,

$$(k-1)V_j + V_t = V_j + V_t = F_{jt} + F_{tj} = (k-1)F_{jt} + F_{tj},$$

and the expression equals zero, as does the coefficient of ϕ_{jt}. It is for this reason that the condition $k \geqslant 3$ was made.

If $v - k = 1$, a test of the interactions can very readily be made, for in this case the parameters, ϕ_{jt}, are orthogonal with the parameters, β_h^*, as well as with $(\gamma_j - \delta_j)$. First, an analysis of variance is worked out ignoring interactions and taking account only of the parameters, α, β_h^* and $(\gamma_j - \delta_j)$; let this give an error sum of squares, S', with

$$[nb(k-1) - (v-1)]$$

degrees of freedom. It is now necessary to evaluate the interaction sum of squares, W, and this is best done by multiplying the estimated value of each interaction parameter, $\hat{\phi}_{jt}$, by the total of the data to which it applies, i.e.

$$W = \sum_{t \neq j} \hat{\phi}_{jt}(T_j - D_{jt}) = -\sum_{t \neq j} \hat{\phi}_{jt} D_{jt}.$$

The quantity S $(= S' - W)$ now becomes the error sum of squares. Since W has $(v^2 - 3v + 1)$ degrees of freedom, S has $[nb(k-1) - v(v-2)]$. A comparison of W and S by the F-test will show if there is any evidence of an interaction of local and remote effects.

If $v-k \geqslant 2$, β_h^* needs to be adjusted to

$$\beta_h^{**} = \beta_h^* - \frac{1}{k}\Sigma'(\phi_{jt}+\phi_{tj}) = \beta_h^* - \frac{1}{k}\Sigma''(\phi_{jt}+\phi_{tj}),$$

where Σ' denotes summation where both treatments j and t are present in block h, and Σ'' denotes similar summation when neither treatment j nor t is present. The choice of summation is a question of convenience. It follows from this that S' will need to be adjusted by adding $Z = \sum_h B_h(\beta_h^* - \beta_h^{**})$, where B_h is the total of data from block h. S now equals $S'+Z-W$, the degrees of freedom for W and S being as before.

Validity of the test for interactions

It is not at once evident that the F-test for interactions is unbiased, but examination shows it to be so. Suppose that there is in fact no interaction, then the parametric expression for y_{hj}, the figure given by treatment j in block h, may be written

$$\alpha + \beta_h^* + (\gamma_j - \delta_j) + x_{hj},$$

where x_{hj} is a random residual. Let T'_j, U'_j, D'_{jt}, etc., be functions of the x's corresponding to T_j, U_j, D_{jt}, etc., as functions of the y's, then $U'_j = E'_{jt} = G' = 0$ and $V'_j = kT'_j$ and $F'_{jt} = kD'_{jt}$. Consequently, if $\hat{\phi}_{jt}$ be evaluated now, i.e. in the case where there are in fact no interactions,

$$\begin{aligned}(v-1)(v-2)(v-k)[(k-1)T'_j+T'_t]-k(v-k)[T'_j+T'_t]\\ -v(v-1)(v-2)[(k-1)D'_{jt}+D'_{tj}]+kv(v-2)[D'_{jt}+D'_{tj}]\end{aligned}$$

will be taken as the value of

$$n(k-2)(v-k)v(v-1)\binom{v-2}{k-2}\hat{\phi}_{jt}.$$

It remains to evaluate W assuming the null-hypothesis. Since $\hat{\phi}_{jt} = \hat{\phi}'_{jt}$, it follows that $W = -\sum_{t \neq j}\hat{\phi}'_{jt}D_{jt}$. Also, D_{jt} equals D'_{jt} plus a linear function of parameters that is determinate for each combination of j and t.

Let ϵ represent expectation under random permutation of plots within blocks, and let $\epsilon(x^2) = \theta$. Since

$$\epsilon(T'_j) = \epsilon(T'_t) = \epsilon(D'_{jt}) = \epsilon(D'_{tj}) = 0,$$

$$\epsilon(T'_jD'_{jt}) = \epsilon(D'^2_{jt}) = n\binom{v-2}{k-1}\theta,$$

$$\epsilon(T'_tD'_{jt}) = \epsilon(D'_{jt}D'_{tj}) = 0,$$

it follows that

$$\epsilon(W) = \frac{k(v^2-3v+1)}{(k-1)}\theta,$$

i.e. $k\theta/(k-1)$ for each degree of freedom, provided the null-hypothesis is true. But this is the same as for S', which is the error sum of squares of a design in balanced incomplete blocks, and hence it must be the same for S, when this equals $S'-W$. Thus, by virtue of the randomization of treatments within blocks, the F-test between W and S is unbiased, at least when $v-1=k$.

In other cases, however, there is the adjustment, Z, to be considered. This has been written as a sum of terms of the form $\phi_{jt}B_h$, where neither treatment j nor t occurs in block h. Since $\epsilon(B'_hT'_j) = \epsilon(B'_hT'_t) = \epsilon(B'_hD'_{jt}) = \epsilon(B'_hD'_{tj}) = 0$, it follows that $\epsilon(Z) = 0$. It thus appears that the F-test between W and S is valid for all the designs considered.

Numerical example

The data of this example are drawn from the first trial to be designed at East Malling using the above approach. As such, it is open to improvement, but is for that reason instructive.

The experiment was intended to study metaxenia in apples. Metaxenia is the phenomenon of pollen affecting the characteristics of the plant pollinated, and it was proposed to test whether it is in fact true that the size of Cox's Orange Pippin apples depends in part on the pollen. In particular, there was a suggestion that pollination by King of the Pippins gives large fruit.

Table 1. *Data of experiment on xenia effects in apples*

Figures represent mean fruit diameters in millimetres

Wave	Block no., h	Kind of pollen, j				Values of B_h
		1	2	3	4	
First	1	—	18·71	17·01	17·23	52·95
	2	20·42	—	18·02	19·59	58·03
	3	21·39	20·85	—	18·94	61·18
	4	19·49	18·48	17·22	—	55·19
Second	5	—	23·16	23·33	23·34	69·83
	6	24·67	—	24·54	25·00	74·21
	7	24·14	21·78	—	20·86	66·78
	8	26·70	25·75	24·77	—	77·22
Values of T_j		136·81	128·73	124·89	124·96	515·39 = G
U_j		122·78	132·24	127·96	132·41	
V_j		+17·82	+3·04	−12·76	−8·10	—

In view of the great labour of emasculating blossoms and then hand-pollinating them, it has previously been the practice to divide the blossom trusses into groups to serve as plots, using whole potted trees as blocks. At the suggestion of the writer, in this trial, conducted in 1952, a balanced incomplete block design was chosen. Pollen from four varieties was to be used ($v = 4$), namely, 1 = King of the Pippins, 2 = Ellison's Orange, 3 = James Grieve and 4 = Worcester Pearmain. The blossoms of each tree were divided into three groups instead of four ($k = 3$), and there were accordingly four kinds of blocks ($b = 4$) depending on the pollen omitted. Twelve trees were selected; four were brought on by heat treatment, four were allowed to develop naturally, and four were retarded by cold, but these last did not appear wholly normal and were discarded. In each of the 'waves' that were used, one tree was allocated at random to make one of each kind of block, and the three pollens applied each to one group of blossoms selected at random. There were thus two blocks of each kind ($n = 2$). The grouping of blocks into waves was not considered above, but gives rise to only trivial modifications of the standard method; it was arranged in order to spread out the labour of applying treatments. Blocks were assigned numbers after randomization, and this accounts for the apparently systematic design.

The figures in Table 1 represent average fruit diameter in millimetres over each group of

blossoms at 49 days after pollination, together with values of B_h, T_j, U_j and V_j. As a check it may be noted that the B's and the T's should sum to G, the U's to $(v-k)G$, and the V's

Table 2. *Values of* $(\gamma_j-\delta_j)$ *and* $24\beta_h^*$

j	1	2	3	4
$(\gamma_j-\delta_j)$	+1·11375	+0·19000	−0·79750	−0·50625

h	1	2	3	4	5	6	7	8
$24\beta_h^*$	−82·88	−49·63	−32·33	−77·92	+52·16	+79·81	+12·47	+98·32

Table 3. *Values of* D_{jt}, F_{jt} *and* $144\hat{\phi}_{jt}$

D_{jt}	$j=1$	$j=2$	$j=3$	$j=4$
$t=1$	—	41·87	40·34	40·57
2	45·09	—	42·56	44·59
3	45·53	42·63	—	39·80
4	46·19	44·23	41·99	—

F_{jt}	$j=1$	$j=2$	$j=3$	$j=4$
$t=1$	—	+2·83	−1·76	−1·07
2	+3·03	—	−4·56	+1·53
3	+8·63	−0·07	—	−8·56
4	+6·16	+0·28	−6·44	—

$144\hat{\phi}_{jt}$	$j=1$	$j=2$	$j=3$	$j=4$
$t=1$	—	+12·90	−19·14	+ 6·24
2	+96·78	—	+ 3·72	−100·50
3	−85·02	− 9·24	—	+ 94·26
4	−11·76	− 3·66	+15·42	—

to zero. Now, $(\gamma_j-\delta_j)$ is estimated by $(v-1)V_j/\{nbk(k-1)\}$, which leads to the values set out in Table 2. Further, B_h equals

$$k\alpha + k\beta_h^* - [\text{sum of } (\gamma_j-\delta_j) \text{ for treatments lacking in block } h].$$

Since $\alpha = G/nbk$, values of $24\beta_h^*$ can be determined as shown in Table 2. Both $(\gamma_j-\delta_j)$ and β_h^* should sum to zero.

From these values it is possible to work out the residual sum of squares within blocks, ignoring interactions. This equals

$$\Sigma\,(\text{data})^2 - G^2/nbk - \sum_j T_j(\gamma_j-\delta_j) - \sum_h B_h\beta_h^*,$$

namely, 6·7350 with 13 degrees of freedom.

To proceed now to the study of possible interactions, the first step is to write down the values of D_{jt} as is done in the top part of Table 3. These should sum horizontally to U_t and vertically to $(v-k)T_j$. The next values needed are those of E_{jt}, but in this instance, where

$v = k+1$, they are all zero and need not be considered further. In the general case they should sum horizontally to $(v-k-1)\,U_t$ and vertically to $(v-k-1)\,U_j$.

From a knowledge of U_t, D_{jt} and E_{jt} it is easy to evaluate F_{jt}; the results of doing so are shown in the middle part of Table 3, and sum horizontally to zero and vertically to $(v-k)\,V_j$.

From these values it is a straightforward matter to obtain estimates of the interaction parameters, $\hat{\phi}_{jt}$, though it is usually easier to estimate some multiple of them. In this example, to maintain generality the full expression has been used, and the figures in the bottom part of Table 3 are of $144\hat{\phi}_{jt}$; had the simplified expression for the case, $v = k+1$, been used, the multiple would have been 48. The figures should sum both ways to zero. Now, $W = -\sum_{t+j} \hat{\phi}_{jt} D_{jt}$, which in this case comes to 3·5340. It has five degrees of freedom. Since $v = k+1$, no adjustment, Z, is needed.

This gives rise to the analysis of variance:

Source	D.F.	S.S.	M.S.
Interactions	5	3·5340	0·7068
Error (by difference)	8	3·2010	0·4001
Total	13	6·7350	

This shows that the interaction effects are not significant; also there is no suggestive regularity in the table of $144\hat{\phi}_{jt}$, so it was decided to ignore possible interactions.

It remains to evaluate γ_j and δ_j separately. This is done quite easily:

$$\gamma_1 = +1{\cdot}754, \quad \delta_1 = +0{\cdot}640,$$
$$\gamma_2 = -0{\cdot}439, \quad \delta_2 = -0{\cdot}629,$$
$$\gamma_3 = -0{\cdot}384, \quad \delta_3 = +0{\cdot}414,$$
$$\gamma_4 = -0{\cdot}931, \quad \delta_4 = -0{\cdot}425.$$

The values of γ_j and δ_j should both sum to zero. The former series is of great interest, because the figures do indeed suggest that pollen from King of the Pippins ($j = 1$) gives larger fruit than do the other pollens. The statistical significance can only be judged from a knowledge of M^0 and M^+, which in turn depend upon M and M'. M' may be evaluated from the analysis of variance just given as 0·4001; to find M it is necessary to consider the values B_h, thus:

	Type of block			
	1	2	3	4
1st wave	52·95	58·03	61·18	55·19
2nd wave	69·83	74·21	66·78	77·22

Eliminating both the effects of type of block and of waves, and remembering that each figure is the sum of three data, these give an error mean square of 7·9216 with three degrees of freedom.

Hence, $M^0 = 8{\cdot}1217$ and $M^+ = 7{\cdot}9716$, and the standard errors of $(\gamma_j - \gamma_k)$ and $(\delta_j - \delta_k)$ are respectively 1·645 and 1·630. Degrees of freedom are very few, but even so the data provide some support for the belief that the kind of pollen affects size of fruit, especially as the effect observed was nominated beforehand. The values of δ_j are not so different as to demonstrate the existence of remote effects.

Discussion

The usefulness of this method depends upon the questions to be answered. If remote effects are to be expected and if the aim is to assess the effect of treating whole plants or animals, nothing is to be gained by treating parts of them. In such a trial the quantities to be deter-

mined are the values of $\alpha+\gamma_j+(k-1)\delta_j$ for the various j's, and the accuracy will in any event depend on M and not at all on M'. Further, if whole organisms are used as plots, the variance of differences between estimated treatment effects will be $2vM/(nbk)$; whereas if they are used as blocks the variance will be $2v(v-1)M/\{nb(v-k)\}$, so the additional labour of using small plots may actually have led to a substantial loss of accuracy.

If, on the other hand, the experimenter is concerned to investigate the mechanism of his treatment effects, the proposed method is of especial value. Thus, in a recent trial the remote effects proved to be nearly equal to the local ones, and this was of some interest, indicating as it did that the local application of a treatment affected all parts of the organism to about the same degree. In other circumstances, it may appear that treatments are strictly local in their effects, or that remote effects exist, and are or are not proportionate to local ones, and if proportionate are of the same or opposite sign. All these phenomena may be suggestive of biological modes of action.

The chief use of the method, however, comes in exploring the desirability of using parts of an organism as plots. If it can be shown that treatments of a certain kind have only local effects, or have remote effects that are proportionate to the local ones, so that they do not invalidate a test, then in future study of such treatments it will be possible to use parts of the same animal or plant as plots without regard to remote effects, and to use any convenient block design. Conclusions will then be based on M' rather than on M with a gain in precision that may be considerable, and less material will be needed. Further, in investigating the possibility of doing this, some information will have been gained as to the effects of the treatments, much of it at a useful level of accuracy.

In using the method described it is clear that the standard error between blocks is as important as that within, and it must, therefore, be based on a sufficient number of degrees of freedom. This was a fault of the trial from which the example was taken, though here it was the original intention to use twelve trees rather than eight. The need to have a reasonable number of degrees of freedom for M means that the method does not lead to a large economy of material as long as local and remote effects are both studied. Of course, if at a later stage it is decided that remote effects can be ignored, and accordingly whole plants or animals can be used as blocks without any of the complications discussed in this paper, the economies may be very great.

SUMMARY

The statistical implications of using organisms as blocks are considered, special attention being given to the case of a treatment applied to one plot having a local effect on the plot to which it is applied and remote effects elsewhere in the block, and also to the case of remote and local effects interacting. A worked example is given, and the usefulness of the method is discussed.

I have been much helped by Mr H. M. Tydeman, who provided the data of the numerical example, and Mr G. H. Freeman, who checked the algebra.

REFERENCE

PATERSON, A. B. & LEECH, F. B. (1954). Factors affecting the intradermal tuberculin reaction on the guinea pig. *Amer. Rev. Tuberc.* **69**, 806–17.

THE USE OF A CONCOMITANT VARIABLE IN SELECTING AN EXPERIMENTAL DESIGN*

By D. R. COX

Department of Biostatistics, School of Public Health, University of North Carolina and Department of Statistics, University of California†

1. Introduction

Suppose that we have t treatments for comparison using N experimental units, for example, animals, and that on each unit we have a preliminary observation, for example, the body weight, available before the treatments are allotted to the units. Suppose that it is expected that the final observation of interest, y, will, in the absence of treatment effects, be well correlated with the preliminary observation, x. Then a number of standard methods are available for exploiting this correlation in order to increase the precision of the estimated treatment effects.

The object of this paper is to compare the methods in some simple situations with a small number of experimental units. Attention is restricted to experiments in which each of the alternative treatments appears the same number, k, of items, so that $N = tk$.

2. Some standard methods

We list first the following methods for increasing the precision of the treatment comparisons.

Method I. An index of response is used, for example, y/x, y/x^2, $y-x$, etc. The index may be suggested by analysis of previous data or by general consideration of a plausible form for the regression relation between y and x.

Method II. The treatments are completely randomized over the experimental units and an adjustment made for regression on x by analysis of covariance. For simplicity only linear regression will be considered.

Method III. The experimental units are ranked in order of increasing x and then grouped into k blocks of t units each, the first block, for example, consisting of the t units with lowest values of x. A randomized block arrangement is then constructed, based on this grouping of the units.

Method IV. This is applicable when k is equal to, or is a multiple of, t. The units are again ranked in order of increasing x and a Latin square arrangement used to control not only differences between blocks but also variation associated with order within a block. Thus with $t = k = 4$, $N = 16$, the units are numbered from 1, ..., 16, in order of increasing x and a Latin square set out, as in the following example:

	Order within block			
Block no.	1	2	3	4
1	$1:T_3$	$2:T_1$	$3:T_2$	$4:T_4$
2	$5:T_4$	$6:T_2$	$7:T_3$	$8:T_1$
3	$9:T_2$	$10:T_4$	$11:T_1$	$12:T_3$
4	$13:T_1$	$14:T_3$	$15:T_4$	$16:T_2$

* This paper was prepared with the partial support of the Air Research and Development Command, under contract with the U.S.A.F. School of Aviation Medicine.

† Present address, Birkbeck College, University of London.

A similar method, employing Youden squares, could be employed if k is not a multiple of t, but would require a lengthier analysis.

Method V. Methods II and III may be combined by using the randomized block design plus a covariance correction.

Method VI. It often happens that the preliminary observation x is not the only thing associated with the units that can be used to increase precision. One of a number of possibilities is that a grouping into randomized blocks should be made on the basis of the other criteria and then a covariance adjustment made for x.

Method VII. A systematic, or non-randomized, design may be used, chosen, for example, to give maximum precision under the hypothesis that the regression equation of y on x is a polynomial of low degree (Cox, 1951). The use of such a design has the disadvantages attendant on the lack of full randomization.

An eighth method is that of Papadakis (see, for example, Bartlett (1938)). This is useful in large experiments in which there are expected to be trends or serial correlation between experimental units adjacent in space or time, but will not be considered here.

3. Basis for comparison

Denote the t treatments by $T_1, \ldots, T_t$ and the corresponding estimated treatment means, after adjustment by covariance where appropriate, by $\hat{y}_1, \ldots, \hat{y}_t$. We measure the random errors associated with the design by the variance of the estimated difference between a pair of treatments, averaged over all pairs of treatments, i.e. by

$$V_t = \operatorname*{Ave}_{i,j,\, i \neq j} V(\hat{y}_i - \hat{y}_j). \tag{1}$$

This is a certain multiple of the population residual variance σ_0^2 and is not affected by errors in estimating σ_0^2. Call (1) the *true* (average) *imprecision* of the design; it is slightly different from the quantity suggested by Lucas and used by Greenberg (1953).

Often, however, we are interested in the *apparent imprecision*, V_a, making due allowance for the effective loss of information that arises in estimating the residual variance. To obtain V_a we use Fisher's factor $(f+3)/(f+1)$, where f is the number of degrees of freedom available to estimate the residual variance (Cochran & Cox, 1950, p. 26); that is, we put

$$V_a = V_t\left(\frac{f+3}{f+1}\right). \tag{2}$$

This will be supplemented by the rather arbitrary rule that if $f < 5$ no effective estimate of the residual variance is considered possible from the data alone.

From the point of view of the length of confidence intervals and the sensitivity of significance tests, a situation with $V_a = 1$ and with the variance estimated from the data is nearly equivalent to a situation in which the residual variance is known and $V_t = 1$.

The use of the average variance in (1) is a natural step but should be viewed critically, particularly in small experiments where there may be substantial variation in precision between different randomization patterns and between different comparisons within one randomization pattern. This is particularly so with Method II.

4. Calculation of imprecision

To compare the different methods it will be assumed that, in the absence of treatment effects, y and x have a bivariate normal frequency distribution with correlation coefficient, ρ, and with the variance of y for fixed x denoted by σ_0^2. If the effect of variation accounted for by x were completely removed, we should have $V_t = 2\sigma_0^2/k$. Therefore we write

$$V_t = \frac{2\sigma_0^2}{k} I_t, \tag{3}$$

$$V_a = \frac{2\sigma_0^2}{k} I_a, \tag{4}$$

and use I_t and I_a as indices of true and apparent imprecision. Clearly $I_a > I_t \geqslant 1$.

We now compute I_t and I_a for the various procedures set out in § 2.

Table 1. *Loss of precision from using wrong index of response*

ρ	Range of β_0/β within which $I_t \leqslant$			I_t if x ignored
	1·1	1·2	1·5	
0·2	(−0·55, 2·55)	(−1·19, 3·19)	(−2·46, 4·46)	1·04
0·4	(0·28, 1·72)	(−0·02, 2·02)	(−0·62, 2·62)	1·19
0·6	(0·68, 1·32)	(0·40, 1·60)	(0·06, 1·94)	1·56
0·8	(0·76, 1·24)	(0·67, 1·33)	(0·47, 1·53)	2·78
0·9	(0·82, 1·18)	(0·74, 1·26)	(0·59, 1·41)	5·26
0·95	(0·90, 1·10)	(0·85, 1·15)	(0·77, 1·23)	10·26

Method I. Suppose that the true regression coefficient of y on x is β and that the index of response is $y - \beta_0 x$. This situation has been considered by Gourlay (1953). It is easily shown that the index of true imprecision is

$$I_t^{(1)} = \begin{cases} \left(1 - 2\rho^2 \dfrac{\beta_0}{\beta} + \rho^2 \dfrac{\beta_0^2}{\beta^2}\right)(1-\rho^2)^{-1} & (\rho \neq 0), \\ 1 + \beta_0^2 \sigma_x^2 / \sigma_y^2 & (\rho = 0). \end{cases} \tag{5}$$

If no attempt is made to use x, i.e. if $\beta_0 = 0$, $I_t = (1-\rho^2)^{-1}$. Thus the attempted correction is an advantage whenever $0 < \beta_0/\beta < 2$. Table 1 shows the ranges of values of β_0/β for which $I_t \leqslant 1{\cdot}1, 1{\cdot}2, 1{\cdot}5$ and so tells us how near β and β_0 need to be to avoid losing a specified amount of information.

Fisher (1935, p. 163) criticized the use of indices of response based on inadequate assessments of the relation between y and x. It is no contradiction of these criticisms to conclude from Table 1 that, particularly if ρ is not near unity, β_0 does not need to be very near β to give a worth-while increase in precision. The use of non-linear indices such as y/x^α allows curvilinear regression to be accounted for. Note, however, that if the treatment effects are constant, independently of x, on the y scale, they will not be constant when the index is used; if the index and the original observation have equal physical significance there may

well be no reason for expecting one rather than the other to be the appropriate scale for measuring treatment effects. However, if the object is the estimation of the effect on the long-run average of y of a change in treatments, there may occasionally be difficulties connected with a naïve use of average of y/x^{α}.

Method II. Denote the terms of the analysis of covariance as follows:

Treatments	T_{xx}	T_{xy}	T_{yy}
Residual	R_{xx}	R_{xy}	R_{yy}
Total	S_{xx}	S_{xy}	S_{yy}

where, for example, R_{xy} denotes the residual sum of products of x and y. If $\bar{x}_i$ is the mean of x for the ith treatment and $\hat{y}_i$ the mean of y adjusted for regression on x,

$$\left.\begin{aligned} V(\hat{y}_i-\hat{y}_j) &= \frac{2\sigma_0^2}{k}\left\{1+\frac{(\bar{x}_i-\bar{x}_j)^2}{2R_{xx}}\right\}, \\ \operatorname*{Ave}_{i,j} V(\hat{y}_i-\hat{y}_j) &= \frac{2\sigma_0^2}{k}\left\{1+\frac{T_{xx}}{(t-1)R_{xx}}\right\}. \end{aligned}\right\} \tag{6}$$

This is conditional on fixed x's; but the x's are in fact a random sample from a normal population and the second term in brackets is therefore proportional to an F variate. Hence if we take expectations and remove the factors $2\sigma_0^2/k$, we obtain for the index of true imprecision

$$\begin{aligned} I_t^{(2)} &= 1+\frac{1}{(k-1)t-2} \\ &= \frac{N-t-1}{N-t-2}. \end{aligned} \tag{7}$$

Also
$$I_a^{(2)} = I_t^{(2)}\frac{N-t+2}{N-t}, \tag{8}$$

since the residual degrees of freedom, after adjusting for regression, are $N-t-1$.

Table 2 gives the values of (7) and (8) for various values of t, k with $N \leqslant 20$.

The increase in variance above that to be expected when the effect of x is completely eliminated can be described as due to errors in estimating the residual regression coefficient, or as arising from the non-orthogonality present in the linear set-up chosen to represent the populations sampled.

Method III. Consider first one block and a fixed set of x values $x_1, \ldots, x_t$. In the absence of treatment effects, the corresponding observations on y are $\beta x_1+\epsilon_1, \ldots, \beta x_t+\epsilon_t$, where β is the regression coefficient of y on x and $\epsilon_1, \ldots, \epsilon_t$ are independent with constant mean and with variance σ_0^2, and determine the dispersion of y about its regression line on x. Hence if two positions are selected randomly to carry treatments T_i and T_j, the expected mean-square difference between the resulting observations is

$$\sigma_0^2+\frac{\beta^2}{t-1}\Sigma(x_i-\bar{x})^2 = \sigma_0^2\left\{1+\frac{\beta^2}{(1-\rho^2)}\frac{1}{(t-1)}\frac{\Sigma(x_i-\bar{x})^2}{\sigma_x^2}\right\}.$$

Hence, averaging over the k blocks, and over the distribution of the x's, we have that

$$I_t^{(3)} = 1+\frac{\rho^2}{1-\rho^2}W, \tag{9}$$

where W is the expected mean square of x within blocks, divided by the variance of x.

Table 2. *Measure of imprecision for various designs*

N	9	12	12	12	15	15	16	16	18	18	18	20	20	20
t	3	4	3	2	5	3	4	2	6	3	2	5	4	2
k	3	3	4	6	3	5	4	8	3	6	9	4	5	10
$I_t^{(2)}$	1·25	1·17	1·14	1·12	1·12	1·10	1·10	1·08	1·10	1·08	1·07	1·08	1·07	1·06
$I_a^{(2)}$	1·67	1·46	1·43	1·35	1·35	1·28	1·28	1·24	1·28	1·22	1·21	1·22	1·21	1·18
W	0·2498	0·2390	0·1715	0·1112	0·2325	0·1278	0·1633	0·0763	0·2282	0·1005	0·0654	0·1584	0·1214	0·0570
$I_t^{(3)}$: $\rho = 0$	1·00	1·00	1·00	1·00	1·00	1·00	1·00	1·00	1·00	1·00	1·00	1·00	1·00	1·00
0·4	1·05	1·05	1·03	1·02	1·04	1·02	1·03	1·01	1·04	1·02	1·01	1·03	1·02	1·01
0·6	1·14	1·13	1·10	1·06	1·13	1·07	1·09	1·04	1·13	1·06	1·04	1·09	1·07	1·03
0·8	1·44	1·42	1·30	1·20	1·41	1·23	1·29	1·14	1·41	1·18	1·12	1·28	1·22	1·10
0·9	2·07	2·02	1·73	1·47	1·99	1·54	1·70	1·33	1·97	1·43	1·28	1·68	1·52	1·24
0·95	3·31	3·21	2·59	2·03	3·15	2·18	2·51	1·71	3·11	1·93	1·61	2·47	2·12	1·53
$I_a^{(3)}/I_t^{(3)}$	*	1·29	1·29	1·33	1·22	1·22	1·20	1·25	1·18	1·18	1·22	1·15	1·15	1·20
W'	0·1015	—	—	—	—	—	0·0676	0·0436	—	0·0540	—	—	—	—
$I_t^{(4)}$: $\rho = 0$	1·00	—	—	—	—	—	1·00	1·00	—	1·00	—	—	—	—
0·4	1·02	—	—	—	—	—	1·01	1·01	—	1·01	—	—	—	—
0·6	1·06	—	—	—	—	—	1·04	1·02	—	1·03	—	—	—	—
0·8	1·18	—	—	—	—	—	1·12	1·08	—	1·10	—	—	—	—
0·9	1·43	—	—	—	—	—	1·29	1·19	—	1·23	—	—	—	—
0·95	1·94	—	—	—	—	—	1·63	1·40	—	1·50	—	—	—	—
$I_a^{(4)}/I_t^{(4)}$	*	—	—	—	—	—	1·29	*	—	1·29	—	—	—	—
$I_t^{(5)}$	1·25	1·17	1·17	1·20	1·12	1·12	1·11	1·14	1·10	1·10	1·12	1·08	1·08	1·11
$I_a^{(5)}$	*	1·56	1·56	*	1·41	1·41	1·36	1·47	1·32	1·32	1·41	1·26	1·26	1·36
$I_t^{(5)\prime}$	1·17	1·12	1·11	1·10	1·10	1·08	1·08	1·07	1·08	1·07	1·06	1·07	1·06	1·05
$I_a^{(5)\prime}$	*	1·50	1·48	*	1·38	1·35	1·32	1·38	1·30	1·28	1·33	1·24	1·24	1·29
$I_t^{(6)}$	1·50	1·25	1·25	1·33	1·17	1·17	1·14	1·20	1·12	1·12	1·17	1·10	1·10	1·14
$I_a^{(6)}$	*	1·67	1·67	1·87	1·46	1·46	1·40	1·50	1·35	1·35	1·46	1·28	1·28	1·40

The superfixes (2), (3), ... refer to the method of design. I_t denotes a measure of true variance, taking no account of errors in estimating the residual variance; I_a denotes a measure of apparent variance taking account of the residual degrees of freedom. $I = 1$ represents complete elimination of the variation associated with x. Cases where there are fewer than 5 residual degrees of freedom are marked by an asterisk. The values given for Method V are lower limits.

We need, in order to calculate W, to consider the following. Take an ordered sample $x_{(1)}, \ldots, x_{(n)}$ from a unit normal population. Divide this into blocks as described above and find the mean square within blocks. Then W is the expected value of this mean square and can be calculated for $N \leqslant 20$ from recently published tables of the second moments of order statistics in a normal sample (Teichroew, 1956); Table 2 gives some numerical values derived in this way.

The general conclusion from the values of $I_t^{(2)}$ and $I_t^{(3)}$ is that Method III is somewhat better than Method II if $\rho < 0{\cdot}6$ and that Method II becomes appreciably better than Method III only when ρ is as large as 0·8 or more. It makes little difference if the comparison is based on I_a instead of on I_t. In larger experiments with moderate t and large k, both methods will be effective in reducing the value of I_t to near unity, except when ρ is very near unity, when the use of Method III will be inadvisable. However, Method III will remain reasonably efficient for any form of smooth regression between y and x, not just for linear regression. If the regression is linear, but the distribution of x is leptokurtic, the randomized block method is likely to be relatively less effective due to the end-blocks having units with widely discrepant values of x.

Method IV. The argument is similar when a Latin square is used, except that W', equal to the expected residual mean square in the two-way array of x's formed from the rows and columns of the square, replaces W. Table 2 gives the value of W', and of $I_t^{(4)}$ in certain cases. If we have r squares, each of size $t \times t$, the residual degrees of freedom are $(t-1)(rt-r-1)$, when the residual within squares and the treatment $\times$ squares terms are combined. This number is small in the cases examined.

The additional precision gained by eliminating 'order within blocks' makes the critical value of ρ at which Methods II and IV are approximately equivalent equal to about 0·8.

Method V. This is the use of x simultaneously for blocking and for covariance correction. There are two possibilities. We may analyse the design as a randomized block with covariance, estimating the regression coefficient from the residual line of the analysis of covariance. Equation (6) applies with R_{xx} still defined as the residual sum of squares, this time in the randomized block analysis. We again require the expectation of T_{xx}/R_{xx}; over the randomization with the x's fixed

$$E(T_{xx}/R_{xx}) \geqslant E(T_{xx})/E(R_{xx}) = \frac{1}{k-1}, \tag{10}$$

with near equality in a large design. Hence

$$I_t^{(5)} \geqslant 1 + \frac{1}{(t-1)(k-1)}, \tag{11}$$

and the right-hand side of (11) will be used as an approximation to $I_t^{(5)}$; the error will be of the order of $[1/(t-1)(k-1)]^2$, as can be shown by the expansion methods of large-sample theory.

The second possibility is to analyse the design as if it were completely randomized. This would be in order if the arrangement into blocks has no effect on the y-values other than that due to correlation with the x's, and if the assumptions of the least squares model can be postulated, i.e. if we use more than pure randomization theory. In this case the argument parallel to (10) leads to

$$I_t^{(5)\prime} \geqslant 1 + \frac{1}{t(k-1)}, \tag{12}$$

again with near-equality in a large experiment.

The numerical values in Table 2 and a direct comparison of the formulae show that if $k > 3$ the lower limit for $I_t^{(5)}$ is greater than $I_t^{(2)}$. Hence, under the conditions postulated, Method V is inferior to Method II. Even with the second method of analysis, i.e. with (12), it seems unlikely that there is appreciable gain in *average* precision over Method II, under the conditions assumed.

Similar conclusions are reached from $I_a^{(5)}$ and $I_a^{(5)\prime}$.

Method VI. In this we consider a randomized block design with grouping based on a criterion separate from x. Quantitative investigation of this, based, for example, on the assumption that x, y and the property determining the grouping have a trivariate normal distribution, has not been attempted. We can, however, deal with two limiting cases. The system of blocking may be identical to that based on x. Equation (11) is then applicable. Or the criterion for grouping may be independent of x, in which case

$$I_t^{(6)} = \frac{(t-1)(k-1)-1}{(t-1)(k-1)-2}. \tag{13}$$

For the smaller values of $(t-1)(k-1)$ among the designs investigated, this is about $1{\cdot}15\, I_t^{(2)}$, showing that in these cases the additional system of blocking should be included only if there is a reasonable prospect of a reduction of 20 % or more in residual variance. For the larger values of $(t-1)(k-1)$, $I_t^{(6)}$ and $I_t^{(2)}$ are very nearly equal.

Method VII. This method is theoretically the most efficient one when the observations on y are built up of a polynomial trend on x plus treatment effect plus random error of constant mean and dispersion. The method is best illustrated by an example: suppose that $N = 9$, $t = k = 3$, that a second degree curve is considered adequate to represent the regression of y on x, and that the values of x are in order

$$-1{\cdot}3, \quad -0{\cdot}9, \quad -0{\cdot}8, \quad -0{\cdot}5, \quad -0{\cdot}2, \quad 0{\cdot}0, \quad 0{\cdot}4, \quad 0{\cdot}7, \quad 1{\cdot}1.$$

The most systematic procedure is to start by forming first and second degree orthogonal polynomials, ξ_1', ξ_2', from these observations. If $m_\alpha = \Sigma x_i^\alpha$, these are

$$\left.\begin{aligned} \xi_1' &= x_i - m_1/m_0, \\ \xi_2' &= x_i^2 - \left(\frac{m_0 m_3 - m_1 m_2}{m_0 m_2 - m_1^2}\right) x_i + \left(\frac{m_1 m_3 - m_2^2}{m_0 m_2 - m_1^2}\right). \end{aligned}\right\} \tag{14}$$

The numerical values are

$$\begin{array}{lccccccccc} \xi_1': & -1{\cdot}13, & -0{\cdot}73, & -0{\cdot}63, & -0{\cdot}33, & -0{\cdot}03, & 0{\cdot}17, & 0{\cdot}57, & 0{\cdot}87, & 1{\cdot}27, \\ \xi_2': & 0{\cdot}89, & 0{\cdot}08, & -0{\cdot}07, & -0{\cdot}40, & -0{\cdot}55, & -0{\cdot}56, & -0{\cdot}32, & 0{\cdot}07, & 0{\cdot}86. \end{array}$$

These are then normalized by dividing by $\sqrt{(\Sigma \xi_i'^2)}$ to give ξ_1 and ξ_2, namely,

$$\begin{array}{lccccccccc} \xi_1: & -0{\cdot}50, & -0{\cdot}33, & -0{\cdot}28, & -0{\cdot}15, & -0{\cdot}01, & 0{\cdot}08, & 0{\cdot}25, & 0{\cdot}39, & 0{\cdot}57, \\ \xi_2: & 0{\cdot}57, & 0{\cdot}05, & -0{\cdot}04, & -0{\cdot}26, & -0{\cdot}35, & -0{\cdot}36, & -0{\cdot}21, & 0{\cdot}04, & 0{\cdot}55. \end{array}$$

We have to select from the nine units, three to receive T_1, etc. Let $\Sigma_1 \xi_1$, $\Sigma_1 \xi_2$, etc., denote the sum over those units receiving T_1 of ξ_1, ξ_2. This gives us six numbers. The treatment arrangement that minimizes the sum of squares of these six numbers is very nearly, or exactly, the most precise arrangement. Trial and error shows this arrangement to be

$$\begin{array}{lccccccccc} x & -1{\cdot}3 & -0{\cdot}9 & -0{\cdot}8 & -0{\cdot}5 & -0{\cdot}2 & 0{\cdot}0 & 0{\cdot}4 & 0{\cdot}7 & 1{\cdot}1 \\ & T_1 & T_3 & T_2 & T_2 & T_3 & T_1 & T_1 & T_3 & T_2 \end{array}$$

(a non-randomized block design) with

	T_1	T_2	T_3
$\Sigma\xi_1$	−0·17	0·14	0·05
$\Sigma\xi_2$	0·00	0·25	−0·26

The sum of squares of these numbers is $S_{\min.} = 0{\cdot}1811$, and the value of I_t is, in general, approximately (Cox, 1951)

$$I_t^{(7)} \simeq \left\{1 - \frac{(t-2)}{k(t-1)} S_{\min.}\right\} \{1 - S_{\min.}/k\}^{-1}, \tag{15}$$

which in this case is 1·03. (The residual degrees of freedom would be 4 if both linear and quadratic trend were removed and 5 if only linear trend is removed.)

In general the systematic search for the optimum arrangement is tedious although it is usually possible to find quickly an arrangement that is nearly the best. If the degree of the polynomial is small compared with k, it will usually be possible to find an arrangement with $I_t^{(7)}$ negligibly greater than 1. The disadvantage of this method is the lack of randomization; the method is of most value in single small experiments.

5. Discussion

In deciding what design to use in a particular case, we should consider

(i) the values for imprecision given above;

(ii) the extent to which departure from assumed conditions is likely to affect (i);

(iii) the importance to be attached to simplicity of design, and analysis;

(iv) the extent to which considerations other than precision are relevant.

The general conclusion from the calculations in § 4 is that the methods based on covariance are preferable to the simpler methods based on blocking only if the correlation coefficient between y and x is at least 0·6, and that under the conditions postulated the systematic design, Method VII, is the most precise. For larger experiments all methods except the first are likely to have I near unity.

The main assumptions in the calculations are the linearity of the regression and the normality of the distribution of x. Non-normality of x should have little effect on the efficiency of covariance analysis, while I for the blocking methods will usually be an increasing function of the kurtosis of the distribution of x. If the regression is non-linear but smooth, blocking methods will remain effective, while covariance methods will not, unless the linear component accounts for most of the regression, or multiple covariance used.

The methods of design are all simple except Method VII. Details of analysis are, of course, simpler for methods not involving covariance.

There are two further considerations. The form of the relation between y and x may be of intrinsic interest, either in helping to understand the experimental material or in giving information useful in the design of further experiments. Also we may suspect that the treatment effects are not independent of x, i.e. that there is a treatment $\times x$ interaction. Such an interaction may give useful insight into the mechanism underlying the treatment effects and may also change any practical recommendations to be made from the experiment. If these considerations are relevant we shall normally prefer to use x quantitatively.

I am grateful to Dr B. G. Greenberg for very helpful discussion.

REFERENCES

BARTLETT, M. S. (1938). The approximate recovery of information from replicated experiments with large blocks. *J. Agric. Sci.* **28**, 418–20.

COCHRAN, W. G. & COX, G. M. (1950). *Experimental Designs.* New York: John Wiley and Son.

COX, D. R. (1951). Some systematic experimental designs. *Biometrika*, **38**, 312–23.

FISHER, R. A. (1935). *Design of Experiments.* Edinburgh: Oliver and Boyd.

GOURLAY, N. (1953). Covariance analysis and its applications in psychological research. *Brit. J. Statist. Psychol.* **6**, 25–34.

GREENBERG, B. G. (1953). Use of covariance and balancing in analytical surveys. *Amer. J. Publ. Hlth*, **43**, 692–9.

TEICHROEW, D. (1956). Tables of expected values of order statistics and products of order statistics for samples of size twenty and less from the normal distribution. *Ann. Math. Statist.* **27**, 410–26.

APPROXIMATE CONFIDENCE LIMITS FOR COMPONENTS OF VARIANCE

BY M. G. BULMER

Unit of Biometry, 6 Keble Road, Oxford

1. STATEMENT OF PROBLEM

Suppose that M_1 and M_2 are independent mean-square variates with f_1 and f_2 degrees of freedom and unknown expected values $(\theta+\sigma^2)$ and σ^2 respectively, where θ and σ^2 are non-negative. In other words, $f_1 M_1/(\theta+\sigma^2)$ and $f_2 M_2/\sigma^2$ are independent chi-square variates with f_1 and f_2 degrees of freedom. The problem considered in this paper is to find confidence limits for θ, the difference between the expected values of the two mean squares; that is to say, we want to find a function of M_1 and M_2, $f(M_1, M_2)$, such that, at least approximately,

$$\Pr[f(M_1, M_2) \leqslant \theta] = \alpha, \tag{1}$$

whatever θ and σ^2 are. The function, f, will of course also depend on the value of α, and we shall be interested in the cases when α is either small or large. If $\alpha = 0{\cdot}05$, for example, the assertion '$\theta \leqslant f(M_1, M_2)$' will be correct 95 % of the time and f will be called the upper 95 % confidence limit for θ; if, on the other hand, $\alpha = 0{\cdot}95$, the assertion '$\theta \geqslant f(M_1, M_2)$' will be correct 95 % of the time and f will be called the lower 95 % confidence limit for θ. The interval between the lower and upper $97\frac{1}{2}$ % limits is a 95% confidence interval for θ.

Such confidence limits will often be useful for estimating the components of variance which arise in the infinite model of the analysis of variance. Approximate solutions of this problem have been given in four recent papers (Bartlett, 1953; Green, 1954; Huitson, 1955; Welch, 1956). A different approach is adopted in this paper which leads to a solution, given in equation (6), at once simple and reasonably accurate. In the second half of the paper, the accuracy of this solution is compared with that of the other approximations which have been suggested.

2. AN APPROXIMATE SOLUTION OF THE PROBLEM

It is reasonable to suppose that the confidence limit, $f(M_1, M_2)$, can be written in the form $M_2 g(F)$, where $F = M_1/M_2$; for if all the observations underlying the analysis are multiplied by a constant, c, then M_1, M_2 and θ are each multiplied by c^2 and the confidence limit should be multiplied by c^2 too. Therefore $f(c^2 M_1, c^2 M_2) = c^2 f(M_1, M_2)$ for any c; the result follows if we put $c^2 = 1/M_2$. We must therefore find a function, g, such that

$$\Pr[M_2 g(F) \leqslant \theta] = \alpha, \tag{2}$$

whatever θ and σ^2 are. But

$$\Pr[M_2 g(F) \leqslant \theta] = \Pr[F \leqslant g^{-1}(\theta/M_2)] = \Pr[M_1 \leqslant M_2 g^{-1}(\theta/M_2)],$$

where g^{-1} is the inverse function of g; it is assumed that $g(F)$ is a continuous and increasing function of F and that g^{-1} therefore exists. Now write $m_1 = M_1/(\theta+\sigma^2)$, $m_2 = M_2/\sigma^2$ and $\rho = \theta/\sigma^2$, so that m_1 and m_2 are standardized mean-square variates with f_1 and f_2 degrees of freedom. Then

$$\Pr[M_1 \leqslant M_2 g^{-1}(\theta/M_2)] = \Pr\left[m_1 \leqslant \frac{m_2 g^{-1}(\rho/m_2)}{(\rho+1)}\right],$$

and we want to find a function g which satisfies the equation

$$\Pr\left[m_1 \leqslant \frac{m_2 g^{-1}(\rho/m_2)}{(\rho+1)}\right] \equiv \int_0^\infty P_1\left[\frac{m_2 g^{-1}(\rho/m_2)}{(\rho+1)}\right] p_2(m_2)\, dm_2 = \alpha \tag{3}$$

for any ρ, where P_1 is the cumulative distribution function of a standardized mean-square variate with f_1 degrees of freedom, and p_2 is the frequency function of a standardized mean-square variate with f_2 degrees of freedom. If this can be done, then $M_2 g(F)$ will be a confidence limit for θ.

Two limiting conditions can be placed on $g(F)$. Let us write L_1 for the lower 100α % point of the F-distribution with f_1 and f_2 degrees of freedom and L_2 for the corresponding lower 100α % point with f_1 and ∞ degrees of freedom.* Then it is obvious that the limit $M_2 g(F)$ should be zero when $F = L_1$. Also $g(F) \sim F/L_2$ as $F \to \infty$, since

$$\Pr[M_2 F/L_2 \leqslant \theta] \equiv \Pr[M_1/\theta \leqslant L_2] \to \alpha \quad \text{as} \quad \rho \to \infty$$

and since F will almost certainly not be very large unless ρ is also very large. Thus we want to find a function, $g(F)$, which satisfies the conditions

$$\left.\begin{aligned} g(L_1) &= 0, \\ g(F) &\sim F/L_2 \quad \text{as} \quad F \to \infty. \end{aligned}\right\} \tag{4}$$

Any function satisfying these conditions will give an exact confidence limit when $\rho = 0$ and when $\rho \to \infty$.

The simplest function satisfying these conditions is the linear approximation, $g_1(F)$:

$$g_1(F) = (F - L_1)/L_2. \tag{5}$$

This is also an exact solution of the problem for all ρ when f_2 is infinite (f_1 remaining finite); for values of ρ between 0 and infinity it gets worse as an approximation the smaller f_2 is. The adequacy of this approximation was tested by direct numerical integration of (3), with g_1^{-1} substituted for g^{-1}, and both the lower and the upper limits were found to be too wide. I therefore introduced a term in F^{-1} and tried as a second approximation

$$g_2(F) = \frac{F}{L_2} - 1 + \frac{L_1}{F}\left(1 - \frac{L_1}{L_2}\right). \tag{6}$$

This approximation satisfies the conditions of (4); the constant -1 was chosen to ensure that it reduces to the linear approximation when f_2 is infinite and the linear approximation is exact. It is shown in the next section that the error $(P - \alpha)$, where $P = \Pr[M_2 g_2(F) \leqslant \theta]$, is of order $(f_1/f_2)^2$. The approximation $M_2 g_2(F)$ is the one considered in the rest of this paper.

When F is small negative limits will sometimes be obtained for θ. It is suggested that in this case the limit should be taken as zero. If this is done the upper limit can never be less than θ when $\theta = 0$ and the information that θ is non-negative has been wasted. This disadvantage is common to all methods which find a confidence interval for θ by means of upper and lower limits with separately fixed significance levels.

* It is important to notice that when we are considering, for example, the *lower* 95 % limit for θ, then $\alpha = 0{\cdot}95$ and L_1 is the upper 5 % point of F with f_1 and f_2 degrees of freedom; but that when we are considering the *upper* 95 % limit for θ, then $\alpha = 0{\cdot}05$ and L_1 is the lower 5 % point of F with f_1 and f_2 degrees of freedom which is the same as the reciprocal of the upper 5 % point of F with f_2 and f_1 degrees of freedom. Similar remarks apply to L_2 with ∞ substituted for f_2.

As an example of this method, consider an experiment quoted in Davies (1947). Six samples of penicillin were each grown in turn on the same set of twenty-four Petri dishes, and the diameters of the zones of inhibition of bacterial growth were measured in millimetres. The mean square for differences between dishes was 4·61 and had 23 degrees of freedom and the error mean square was 0·304 with 115 degrees of freedom. If we can regard the dishes as a random sample from a normal universe of dishes with variance σ_D^2, then

$$\theta = E(M_1) - E(M_2) = 6\sigma_D^2.$$

The lower and upper 97·5 % confidence limits for θ given by $M_2 g_2(F)$ are 2·48 and 8·80 mm.^{2} respectively; thus the 95 % confidence interval for σ_D^2 is 0·413 to 0·918 mm.^{2}.

3. The accuracy of the approximation

$P = \Pr[M_2 g_2(F) \leqslant \theta]$ is exactly equal to α when ρ is zero or infinite or when f_2 is infinite (f_1 being finite). The purpose of this section is to study the behaviour of P for intermediate values of ρ. If we write $r = f_1/f_2$, then we can show that, for large f_1 and f_2, P depends only on ρ and r and not on the actual values of f_1 and f_2. In fact, if we divide $M_2 g_2(F) - \theta$ by σ^2 we get $P = \Pr[z \leqslant 0]$, where

$$z = \frac{m_1(\rho+1)}{L_2} - m_2 + \frac{m_2^2 L_1}{m_1(\rho+1)}\left(1 - \frac{L_1}{L_2}\right) - \rho. \tag{7}$$

If we let f_1 and f_2 tend to infinity while $r = f_1/f_2$ remains constant, then we can write

$$L_1 = 1 - \xi_\alpha(2/f_1 + 2/f_2)^{\frac{1}{2}} \quad \text{and} \quad L_2 = 1 - \xi_\alpha(2/f_1)^{\frac{1}{2}},$$

where ξ_α is the upper 100α % point of the normal distribution; when $\alpha = 0{\cdot}05$, for example, $\xi_\alpha = +1{\cdot}6449$, and when $\alpha = 0{\cdot}95$, $\xi_\alpha = -1{\cdot}6449$. We can therefore see that z becomes approximately normally distributed with mean

$$\xi_\alpha\left[(\rho+1)\left(\frac{2}{f_1}\right)^{\frac{1}{2}} + \frac{(2/f_1 + 2/f_2)^{\frac{1}{2}} - (2/f_1)^{\frac{1}{2}}}{\rho+1}\right], \tag{8}$$

and variance $(\rho+1)^2(2/f_1) + (2/f_2)$. Thus

$$\frac{E(z)}{\sigma_z} \to \xi_\alpha\left[1 + \frac{(r+1)^{\frac{1}{2}} - 1}{(\rho+1)^2}\right]\left[1 + \frac{r}{(\rho+1)^2}\right]^{-\frac{1}{2}}. \tag{9}$$

For fixed r, this quantity is at its maximum when $(\rho+1)^2 = (1+r)^{\frac{1}{2}} + 1$. Thus the extreme value of P is about

$$\alpha + \xi_\alpha f(\xi_\alpha)\left[1 - \frac{2\{(1+r)^{\frac{1}{2}} - (1+r)^{\frac{1}{4}}\}}{r}\right], \tag{10}$$

where $f(\xi_\alpha)$ is the ordinate of the normal curve at ξ_α. Thus we can calculate that when f_1 and f_2 are large and $\alpha = 0{\cdot}05$, P lies between 0·05 and 0·05252 when $r = 1$; the corresponding maxima for $r = \frac{1}{2}$ and $\frac{1}{4}$ are 0·05087 and 0·05026. It will be noted that $(P - \alpha)$ is of order r^2.

Equation (10) gives us the extreme values of P when f_1 is large. In order to investigate what happens when f_1 is small I have expanded P in the form $\alpha + \alpha_1/f_2 + \alpha_2/f_2^2 + \ldots$. It is shown in the Appendix that $\alpha_1 = 0$ for all ρ. It is also shown how α_2 can be calculated. The maximum and minimum values of $c = \alpha_2/f_1^2$ (multiplied by 100), taken over all possible values of ρ, are given in Table 1, for several values of f_1 and α. This table gives a good idea of the behaviour of the extreme values of P for all values of f_1. When f_1 is large, formula (10)

can be expanded to give $\max c$ (for $\alpha < \frac{1}{2}$) or $\min c$ (for $\alpha > \frac{1}{2}$); in either case the other extremum is zero. In the case $\alpha < \frac{1}{2}$, $\min c$ is zero for all f_1 and $\max c$ increases steadily to its limiting value as f_1 increases. In the case $\alpha > \frac{1}{2}$, the position when f_1 is small is the reverse of that when f_1 is large; $\min c$ is zero and $\max c$ is positive (and is larger the smaller f_1 is). For moderate values of f_1 both the maximum and minimum values of c are non-zero.

Table 1. *The minimum and maximum values* 100c *over all values of* ρ. Min 100c *for* $\alpha < \frac{1}{2}$ *is always zero*

f_1	α	Max 100c	α	Min 100c	Max 100c
2	0·01	0·00	0·99	0	4·38
4	·01	·01	·99	0	1·77
8	·01	·06	·99	0	0·70
12	·01	·09	·99	0	0·40
24	·01	·14	·99	−0·05	0·13
60	·01	·17	·99	−0·12	0·02
120	·01	·19	·99	−0·15	0
∞	·01	·19	·99	−0·19	0
2	0·025	0·00	0·975	0	5·84
4	·025	·02	·975	0	2·38
8	·025	·11	·975	0	0·93
12	·025	·17	·975	−0·02	0·51
24	·025	·25	·975	−0·13	0·15
60	·025	·31	·975	−0·26	0·01
120	·025	·34	·975	−0·30	0
∞	·025	·36	·975	−0·36	0
2	0·05	0·00	0·95	0	6·50
4	·05	·02	·95	0	2·60
8	·05	·16	·95	0	1·01
12	·05	·25	·95	−0·08	0·53
24	·05	·37	·95	−0·27	0·14
60	·05	·46	·95	−0·42	0·01
120	·05	·49	·95	−0·47	0
∞	·05	·53	·95	−0·53	0

If we write c_0 for one of the extreme values of c in Table 1, then the Taylor expansion for the extreme value of P would be $\alpha + c_0 r^2 + \ldots$. However, as $f_1 \to \infty$, $c_0 r^2 \to \xi_\alpha f(\xi_\alpha) r^2/32$ (see equation (A 10) of the Appendix), whereas the extreme value of P tends to (10). We might therefore hope to get a better approximation for the extreme value of P from the equation

$$\alpha + 32c_0\left[1 - 2\frac{\{(1+r)^{\frac{3}{4}} - (1+r)^{\frac{1}{4}}\}}{r}\right], \tag{11}$$

which reduces to (10) when f_1 and f_2 tend to infinity together. The extreme values of P have been obtained by direct numerical integration of (3) in some cases, and these exact values are compared with the approximations given by (11) in Table 2. The approximation seems to be fairly good. It can be concluded that P will not differ appreciably from α in any case

likely to be met with in practice, and that for any value of f_1 the accuracy becomes better rapidly as r becomes smaller (roughly in proportion to r^2).

Table 2. *Comparison of the extreme values of P obtained by numerical integration with the approximate values given by equation* (11)

f_1	f_2	α	Exact range	Approximate range
8	6	0·05	0·0500 – 0·0514	0·0500 – 0·0511
8	12	·05	·05000 – ·05048	·05000 – ·05041
24	24	·05	·0500 – ·0520	·0500 – ·0518
24	48	·05	·05000 – ·05064	·05000 – ·05061
2	6	·95	·9500 – ·9550	·9500 – ·9554
2	12	·95	·9500 – ·9515	·9500 – ·9515
8	6	·95	·9500 – ·9586	·9500 – ·9571
8	12	·95	·9500 – ·9530	·9500 – ·9526
24	24	·95	·9491 – ·9511	·9487 – ·9507
24	48	·95	·94962 – ·95030	·94956 – ·95023

4. COMPARISON WITH OTHER APPROXIMATIONS

$(M_1 - M_2)$ is an unbiased estimator of θ with variance $2w_1(\theta+\sigma^2)^2 + 2w_2\sigma^4$, where $w_1 = 1/f_1$ and $w_2 = 1/f_2$. If we estimate $(\theta+\sigma^2)^2$ by M_1^2 and σ^4 by M_2^2 and regard $(M_1 - M_2)$ as an approximately normal variate, then an approximate confidence limit for θ is

$$M_1 - M_2 + \xi_\alpha(2w_1M_1^2 + 2w_2M_2^2)^{\frac{1}{2}} = M_2[F - 1 + \xi_\alpha(2w_1F^2 + 2w_2)^{\frac{1}{2}}], \qquad (12)$$

where ξ_α is the upper $100\alpha\,\%$ point of the normal distribution. This approximation is well known and will be called the *first normal approximation*. Welch (1956) has developed a series solution, analogous to the Cornish–Fisher expansion, for the general problem of finding confidence limits for linear combinations of several variances. For the special case of the difference between two variances considered in this paper, his first approximation is the first normal approximation given by (12); his second approximation, which will be called the *second normal approximation*, has an error of order (w_1, w_2) and is

$$M_2\left[F - 1 + \xi_\alpha(2w_1F^2 + 2w_2)^{\frac{1}{2}} + \tfrac{2}{3}(2\xi_\alpha^2 + 1)\frac{(w_1^2F^3 - w_2^2)}{(w_1F^2 + w_2)}\right], \qquad (13)$$

and his third approximation, which will be called the *third normal approximation*, has an error of order $(w_1^{\frac{3}{2}}, w_2^{\frac{3}{2}})$ and is

$$M_2\left\{F - 1 + \xi_\alpha(2w_1F^2 + 2w_2)^{\frac{1}{2}} + \tfrac{2}{3}(2\xi_\alpha^2 + 1)\frac{(w_1^2F^3 - w_2^2)}{(w_1F^2 + w_2)} + 2(2w_1F^2 + 2w_2)^{-\frac{1}{2}}\right.$$
$$\left.\times\left[-\xi_\alpha(w_1^2F^2 + w_2^2) + \tfrac{1}{2}(5\xi_\alpha^3 + 9\xi_\alpha)\frac{(w_1^3F^4 + w_2^3)}{(w_1F^2 + w_2)} - \tfrac{1}{9}(16\xi_\alpha^3 + 23\xi_\alpha)\frac{(w_1^2F^3 - w_2^2)^2}{(w_1F^2 + w_2)^2}\right]\right\}. \qquad (14)$$

The second normal approximation is equivalent to Bartlett's (1953) approximation to the order of magnitude to which it is correct.

Huitson (1955) has developed an alternative expansion for the general problem of finding confidence limits for linear combinations of variances. Welch (1956) has also derived this alternative expansion by a slightly different approach. Huitson was mainly concerned with

the problem of estimating the total variability (that is, the *sum* of two or more variances), and his method seems entirely satisfactory for this purposes. It is not satisfactory, however, in the case of the *difference* between two variances considered in this paper. Huitson's expansion, when applied to this special case, becomes

$$\frac{M_2(F-1)}{1-\xi_\alpha(2w_1F^2+2w_2)^{\frac{1}{2}}(F-1)^{-1}+\ldots}. \tag{15}$$

It is easily seen that this is an unsuitable approximation near $F = 1$. For consider (15) taken as far as the large sample approximation (i.e. as far as it is actually written explicitly). The limit is zero only when $F = 1$, which contradicts the first condition of (4). Furthermore, the upper limit is infinite for some critical value of F, say F_U, greater than 1 for which the denominator becomes zero and the lower limit becomes infinite for another value of F, say F_L, less than 1. When $F_L < F < F_U$, the upper limit is negative and the lower limit is positive! It is true that (15) leads to Welch's first expansion if the denominator $(1-h)^{-1}$ is expanded in a formal Taylor series; but, when F is near 1, h becomes very large and the Taylor expansion diverges.

Table 3. *The limiting probabilities that θ is greater than the three normal approximations as either f_2 or ρ approaches infinity*

f_1	First normal approximation		Second normal approximation		Third normal approximation	
	$\alpha=0{\cdot}05$	$\alpha=0{\cdot}95$	$\alpha=0{\cdot}05$	$\alpha=0{\cdot}95$	$\alpha=0{\cdot}05$	$\alpha=0{\cdot}95$
30	0·11392	0·99266	0·06303	0·92551	0·05306	0·95550
60	·09412	·98377	·05688	·93931	·05120	·95185
120	·08051	·97544	·05359	·94513	·05045	·95063
240	·07121	·96870	·05186	·94770	·05017	·95022
480	·06480	·96354	·05095	·94889	·05006	·95007
960	·06035	·95973	·05049	·94946	·05002	·95003

The accuracy of the three normal approximations (12), (13) and (14) depends on both f_1 and f_2 being large; the accuracy of my approximation depends mostly on the ratio f_1/f_2 and does not get any better if they are both increased together. We can investigate the accuracy of the normal approximations as either f_2 or ρ approaches infinity as follows. When $f_2 \to \infty$, the first normal approximation (12), for example, becomes $M_1 - \sigma^2 + \xi_\alpha M_1(2w_1)^{\frac{1}{2}}$ and

$$P = \Pr[M_1 + \xi_\alpha M_1(2w_1)^{\frac{1}{2}} \leqslant \theta + \sigma^2] = \Pr[m_1 \leqslant (1+\xi_\alpha(2w_1)^{\frac{1}{2}})^{-1}],$$

where m_1 is a standardized mean-square variate with f_1 degrees of freedom. When f_2 is finite but $\rho \to \infty$, terms in (12) not involving F can be ignored and

$$P = \Pr[M_1 + \xi_\alpha M_1(2w_1)^{\frac{1}{2}} \leqslant \theta] = \Pr[m_1 \leqslant (1+\xi_\alpha(2w_1)^{\frac{1}{2}})^{-1}],$$

as before. Thus the limiting probability as either f_2 or ρ tends to infinity can be obtained from tables of the chi-square integral. The limiting values for the second and third normal

approximations corresponding to the reciprocal of $1+\xi_\alpha(2w_1)^{\frac{1}{2}}$ are easily found to be the reciprocals of

$$1+\xi_\alpha(2w_1)^{\frac{1}{2}}+\tfrac{2}{3}(2\xi_\alpha^2+1)w_1 \quad \text{and} \quad 1+\xi_\alpha(2w_1)^{\frac{1}{2}}+\tfrac{2}{3}(2\xi_\alpha^2+1)w_1+\tfrac{1}{36}(2w_1)^{\frac{3}{2}}(13\xi_\alpha^3+17\xi_\alpha)$$

respectively. The corresponding probabilities for $\alpha = 0{\cdot}05$ and $0{\cdot}95$ are given in Table 3. We shall certainly not underrate the accuracy of the approximations if we confine ourselves to these special cases. Thus the first normal approximation is rather bad even when f_1 is as large as 1000; the second normal approximation will only be better than mine under rather exceptional circumstances (the maximum deviations $|P-\alpha|$ for my approximation are, when f_1 is large and $\alpha = 0{\cdot}05$ or $0{\cdot}95$, about $0{\cdot}00252$, $0{\cdot}00087$ or $0{\cdot}00026$ when f_1/f_2 is 1, $\frac{1}{2}$ or $\frac{1}{4}$). The third normal approximation will often be better than mine, but it seems doubtful whether the gain in accuracy is worth the increased labour of evaluating it.

The accuracy of Huitson's approximation (15) for large ρ can be investigated in the same way, although the argument about the limiting probability as $f_2 \to \infty$ no longer holds. It turns out that if we take Huitson's approximation as far as the term of order $(w_1^{\frac{1}{2}i}, w_2^{\frac{1}{2}i})$, then

Table 4. *The limiting probabilities that θ is greater than Huitson's approximation as ρ approaches infinity*

f_1	Huitson's first approximation		Huitson's second approximation		Huitson's third approximation	
	$\alpha = 0{\cdot}05$	$\alpha = 0{\cdot}95$	$\alpha = 0{\cdot}05$	$\alpha = 0{\cdot}95$	$\alpha = 0{\cdot}05$	$\alpha = 0{\cdot}95$
30	0·03066	0·93826	0·04823	0·95107	0·05009	0·95004
60	·03725	·94105	·04919	·95058	·05002	·95001
120	·04145	·94334	·04962	·95031	·05000	·95000
240	·04416	·94511	·04982	·95016	·05000	·95000
480	·04598	·94645	·04991	·95008	·05000	·95000
960	·04720	·94744	·04995	·95004	·05000	·95000

as $\rho \to \infty$, $P \to \Pr[m_1 \leqslant C]$ for all f_2, where C is the Cornish–Fisher expansion for the percentage point of m_1 taken as far as the corresponding term. These limiting probabilities are given in Table 4. It can be seen that for $\rho \to \infty$ Huitson's approximation is considerably better than the normal approximation of the corresponding order. It might, therefore, be possible to obtain a much better approximation by using the normal approximation for small F and Huitson's approximation for large F.

Daniels (1939) has proposed a slight modification of the first normal approximation (12). For he points out that M_1^2 and M_2^2 are not unbiased estimators of $(\theta+\sigma^2)^2$ and σ^4, whereas $f_1 M_1^2(f_1+2)^{-1}$ and $f_2 M_2^2(f_2+2)^{-1}$ are; he suggests that the sampling variance of (M_1-M_2) should be estimated by $2M_1^2(f_1+2)^{-1}+2M_2^2(f_2+2)^{-1}$ instead of by $2M_1^2 f_1^{-1}+2M_2^2 f_2^{-1}$. The first normal approximation is, however, so bad that Daniels's modification makes little difference.

Green (1954) has developed approximate confidence limits by a method more like the one adopted in this paper. His solution, however, is not explicit and is incapable of application until it has been tabulated. Green gives only one test of the accuracy of his approximation,

that $P = 0{\cdot}97492$ when $f_1 = 8, f_2 = 50, \rho = 1$ and $\alpha = 0{\cdot}975$. From equation (11) it can be calculated that P for my approximation varies between 0·97500 and 0·97520 when $f_1 = 8$, $f_2 = 50$ and $\alpha = 0{\cdot}975$.

APPENDIX

We wish to expand

$$P = \int_0^\infty P_1\left[\frac{m_2 g_2^{-1}(\rho/m_2)}{\rho+1}\right] p_2(m_2)\,dm_2 \tag{A 1}$$

in the form $P = \alpha + \alpha_1/f_2 + \alpha_2/f_2^2 + \ldots$, and in particular to show that $\alpha_1 = 0$.

Now, solving equation (6) for F in terms of $g_2(F) = v$, say, we find

$$F = g_2^{-1}(v) = \tfrac{1}{2}L_2(1+v) + [\tfrac{1}{4}L_2^2(1+v)^2 - L_1(L_2 - L_1)]^{\frac{1}{2}}. \tag{A 2}$$

(The positive square root must be taken since $v = g_2(F) \sim F/L_2$ for large F, or $F \sim L_2 v$ for large v.) Expanding (A 2) in powers of $1/f_2$ we have

$$g_2^{-1}(v) = L(1+v) + \frac{L'}{f_2(1+v)} + \frac{\tfrac{1}{2}L''(1+v)^2 + (L'^2/L)(2v+v^2)}{f_2^2(1+v)^3} + O(f_2^{-}), \tag{A 3}$$

where $L = L_2$, L', L'' denote the successive derivatives of L_1 with respect to f_2^{-1} at $f_2 = \infty$. Writing $m_2 = 1 + x$ (so that the expected value of x is zero) we now find, expanding in powers of f_2^{-1} and x,

$$\begin{aligned}\frac{m_2 g_2^{-1}(\rho/m_2)}{\rho+1} = L + \frac{L}{1+\rho}x + \frac{L'}{f_2}\left[\frac{1}{(1+\rho)^2} + \frac{1+2\rho}{(1+\rho)^3}x + \frac{\rho^2 x^2}{(1+\rho)^4}\right] \\ + \frac{1}{f_2^2}\left[\frac{L''}{2(1+\rho)^2} + \frac{L'^2(2\rho+\rho^2)}{L(1+\rho)^4}\right] + O(f_2^{-3}, f_2^{-2}x, f_2^{-1}x^3),\end{aligned} \tag{A 4}$$

and, expanding P_1 about L, we find

$$\begin{aligned}P_1\left[\frac{m_2 g_2^{-1}(\rho/m_2)}{\rho+1}\right] = \alpha + \frac{pL}{1+\rho}x + \frac{p'L^2}{2(1+\rho)^2}x^2 + \frac{p''L^3}{6(1+\rho)^3}x^3 + \frac{p'''L^4}{24(1+\rho)^4}x^4 \\ + \frac{1}{f_2}\left[\frac{pL'}{(1+\rho)^2} + \frac{pL'(1+2\rho)+p'LL'}{(1+\rho)^3}x + \frac{pL'\rho^2+p'LL'(1+2\rho)+\frac{1}{2}p''L^2L'}{(1+\rho)^4}x^2\right] \\ + \frac{1}{f_2^2}\left[\frac{pL''}{2(1+\rho)^2} + \frac{pL'^2(2\rho+\rho^2)/L+\frac{1}{2}p'L'^2}{(1+\rho)^4}\right] + O(f_2^{-3}, f_2^{-2}x, f_2^{-1}x^3, x^5).\end{aligned} \tag{A 5}$$

Here $p, p', \ldots,$ denote $p_1(L) = P_1'(L), p_1'(L), \ldots,$ and we note that $P_1(L) = \alpha$. Substituting this expression in (A 1), and noting that $\int x^r p_2(m_2)\,dm_2$ is the rth central moment of m_2, we find

$$\begin{aligned}P = \alpha + \frac{p'L^2+pL'}{f_2(1+\rho)^2} + \frac{1}{f_2^2(1+\rho)^4}[\tfrac{1}{8}p''L^2(1+\rho) + \tfrac{1}{2}p'''L^4 + 2pL'\rho^2 \\ + 2p'LL'(1+2\rho) + p''L^2L' + \tfrac{1}{2}pL''(1+\rho)^2 + (pL'^2/L)(2\rho+\rho^2) + \tfrac{1}{2}p'L'^2] + O(f_2^{-3}).\end{aligned} \tag{A 6}$$

This expression can be thrown into a more convenient form by replacing the derivatives of the percentage point $L, L', \ldots$ by their expressions in terms of $p, p', \ldots$. These relations can be obtained by expanding the expression

$$\int_0^\infty P_1(L_1 m_2)\, p_2(m_2)\,dm_2 = \alpha \tag{A 7}$$

in powers of $1/f_2$ and equating coefficients. The same result can be found simply by putting $\rho = 0$ in (A 6) (when the left-hand side becomes exactly α). Thus, as remarked previously, the coefficient of $1/f_2$ in (A 6) vanishes identically, and we have finally

$$\begin{aligned}P = \alpha + \frac{1}{f_2^2(1+\rho)^4}\left[-\tfrac{1}{8}p''L^2(\rho+\rho^2) - 2p'L^2\rho^2 + \frac{p'^2L^3}{p}(2\rho+3\rho^2)\right. \\ \left. + \left(-\tfrac{1}{2}p'''L^4 + \frac{p'p''L^4}{p} - \frac{\frac{1}{2}p'^3L^4}{p^2}\right)(2\rho+\rho^2)\right] + O(f_2^{-3}).\end{aligned} \tag{A 8}$$

The quantity c defined in § 3 is the second term on the right of (A 8) with the factor $1/f_2^2$ replaced by $1/f_1^2$. It will be recollected that $L = L_2$ is the lower 100α % point of an F distribution with f_1 and

∞ degrees of freedom, or of a χ^2/f_1 distribution with f_1 degrees of freedom; $p, p', \ldots$ represent the ordinate of this distribution at the point $\chi^2/f_1 = L$, and its derivatives with respect to L. The latter quantities can be conveniently evaluated using Molina's (1945) tables of the Poisson distribution.

Now consider the case where f_1 is large. Then the above χ^2/f_1 distribution becomes nearly normal, with mean 1 and variance $2/f_1$, so that L is approximately $1-\xi_\alpha(2/f_1)^{\frac{1}{2}}$, where ξ_α is the upper 100α % normal deviate. If we write $f(\xi_\alpha)$ for the ordinate of the standardized normal curve at ξ_α, we have approximately

$$p = (\tfrac{1}{2}f_1)^{\frac{1}{2}} f(\xi_\alpha), \quad p^{(r)} = (\tfrac{1}{2}f_1)^{\frac{1}{2}r+\frac{1}{2}} H_r(\xi_\alpha) f(\xi_\alpha),$$

where the H_r are Hermite polynomials. Making these substitutions in the expression for c provided by (A 8), we easily find

$$c = \frac{1}{8}\frac{2\rho+\rho^2}{(1+\rho)^4}\xi_\alpha f(\xi_\alpha) + O(f_1^{-\frac{1}{2}}), \tag{A 9}$$

an expression also obtainable from (9). This function has a maximum at $\rho = \sqrt{2}-1$, when it takes the value

$$\max c = \tfrac{1}{32}\xi_\alpha f(\xi_\alpha) \quad (f_1 \text{ large}), \tag{A 10}$$

as indicated at the end of § 3. This result can also be obtained from equation (10).

I should like to thank Mr A. M. Walker for helpful discussions; I should also like to thank the referee for a valuable criticism of the first draft of this paper.

REFERENCES

BARTLETT, M. S. (1953). *Biometrika*, **40**, 306–17.
DANIELS, H. E. (1939). *Suppl. J.R. Statist. Soc.* **6**, 186–97.
DAVIES, O. L. (1947). *Statistical Methods in Research and Production*. London: Oliver and Boyd.
GREEN, J. R. (1954). *Ann. Math. Statist.* **25**, 671–86.
HUITSON, A. (1955). *Biometrika*, **42**, 471–9.
MOLINA, E. C. (1945). *Poisson's Exponential Binomial Limit*. New York: Van Nostrand.
WELCH, B. L. (1956). *J. Amer. Statist. Ass.* **51**, 132–48.

Editorial Note. The accuracy of Green's solution has been further investigated in a Thesis entitled 'A confidence interval for variance components' by S. Lipton, presented for the Degree of M.Sc. of Liverpool University in 1955.

MULTIPLE RUNS

By D. E. BARTON and F. N. DAVID
University College, London

1. The number of ways in which r_1 elements of one kind and r_2 of another can be arranged in a line to form a sequence of $2t$ or $2t+1$ groups was solved by Whitworth (1886, Problems 193 and 194), and it was probably not new to him. The solution was revived by Stevens (1939), Wald & Wolfowitz (1940) and Mood (1940). Sequences in which there are k different kinds of elements have also been treated. Whitworth takes three kinds of elements with the same number of each while Mood obtains a general solution for the distribution of the number of runs of one kind, given k types of element. Using a simple generalization of Whitworth's method we show here how the distributions of the total number of runs can be built up for the multiple case from that for two alternatives. The method of the characteristic random variable is used to obtain reasonably compact expressions for the moments of the total number of multiple runs. The assumption of normality for this total number is not discussed by us at length, since it obviously follows from the work of Wald & Wolfowitz and of Mood. We show, however, an alternative limit which is Poisson.

2. Starting with the two alternatives case it is assumed that there are r_1 white balls and r_2 red balls. If they are arranged in a line randomly, then T, the total number of groups (or runs) will be made up of t white groups $+t$ red groups, or $(t+1)$ white and t red, or t white and $(t+1)$ red. The number in the fundamental probability set is

$$ {}^{r}C_{r_1}, $$

where $$ r = r_1 + r_2. $$

The probability distribution of T is

$$ P\{T = 2t\} = 2\,{}^{r_1-1}C_{t-1}\,{}^{r_2-1}C_{t-1}/{}^{r}C_{r_1} $$

and $$ P\{T = 2t+1\} = P\{T = 2t\}\left(\frac{r-2t}{2t}\right). $$

Now let there be r_1 white, r_2 red and r_3 black balls, and suppose it is required to find the probability distribution of

$$ T = t_1 + t_2 + t_3, $$

where t_1 is the number of white groups, t_2 the number of red and t_3 the number of black groups, respectively. The number of ways in which r_i $(i = 1, 2, 3)$ can be split into t_i groups is

$$ {}^{r_i-1}C_{t_i-1}, $$

so the total number of ways is $$ \prod_{i=1}^{3} {}^{r_i-1}C_{t_i-1}. $$

We now consider the t_i as units and look for the number of ways in which they can be arranged along a line so that no two like colours are together. Following Whitworth we take the t_2 and t_3 groups and put them down in any order. Suppose that there is a total of x contacts of the same colour (RR or BB). There will be t_2+t_3-1 total contacts so that the number of RB or BR contacts will be t_2+t_3-1-x. Now take up the t_1 white groups. x of

them will have to go between the RR and BB contacts leaving $t_1 - x$ to be placed between the BR and RB contacts or at either end of the line. This can be done in

$$t_2+t_3+1-xC_{t_1-x}$$

ways. The number of ways in which t_2 and t_3 can be arranged to produce t_2+t_3-1-x contacts is the number of ways in which t_2 and t_3 can be arranged to produce $t_2+t_3-x = G$ (say) groups. The number of ways for this, depending on whether G is even or odd, is quoted above. The total number of ways of producing

$$T = t_1+t_2+t_3$$

will be then $$\sum \prod_{i=1}^{3} r_i-1C_{t_i-1} \sum_x t_2+t_3+1-xC_{t_1-x}[P\{G = t_2+t_3-x\}\, t_2+t_3C_{t_2}],$$

where the inner sum is taken over all possible values of x and the outer sum is over all possible 3-partitions of T. There are restrictions on both x and T in order that the combinatorial coefficient shall have sense, but these are obvious in any calculation. The number in the fundamental probability set is

$$\frac{r!}{\prod_{i=1}^{3} r_i!},$$

where r is the sum of the r_i. The probability of obtaining a given number of runs is just the ratio of the two expressions.

3. The distributions for four and more runs can be built up successively from the distribution for three. Let there be $r_i (i = 1, 2, \ldots, k)$ with r the sum of the r_i. Let there be $t_i (i = 1, 2, \ldots, k)$ groups respectively and suppose that

$$p(T) = p\left(\sum_{i=1}^{k} t_i\right)$$

is required. The number of ways of forming T groups is

$$\prod_{i=1}^{k} r_i-1C_{t_i-1}.$$

Arrange $$T' = \sum_{i=1}^{k-1} t_i = T - t_k$$

in any random order and let there be x total contacts of self-colours, i.e. x is the total number of white-white, red-red, black-black, ..., contacts. Since there are $T' - 1$ contacts in all there will be $T' - 1 - x$ contacts of different colours. Take the t_k groups which so far have not been arranged. Put x of these in the x self-colour contacts and arrange the remaining $t_k - x$ in the $T' - 1 - x$ different-colour contacts, or at the ends of the line. The number of ways of doing this will be

$$T'+1-xC_{t_k-x}.$$

The number of ways in which T' elements can be arranged to form $T' - x$ groups is obtained from the distribution of runs of $k-1$ colours and we have

$$P\{T = T_0\} = \frac{\prod_{i=1}^{k} r_i!}{r!} \sum_{T_0} \prod_{i=1}^{k} r_i-1C_{t_i-1} \sum_x T'+1-xC_{t_k-x} \left[P\left(T' = \sum_1^{k-1} t_i - x\right) \frac{\left(\sum_{j=1}^{k-1} t_j\right)!}{\prod_{j=1}^{k-1} t_j!} \right],$$

where the first sum is over all possible k-partitions to T_0 and the second sum over all possible x. The restrictions on both x and the partitions of T_0 become obvious in calculation and the distributions can be calculated surprisingly quickly. Using the formulae of this and of the preceding section the distributions for $r = 3, \ldots, 12$ and $k = 3, 4$ were found and are given in Tables 1 and 2. More extensive calculations did not appear profitable in the light of approximations discussed later.

4. The moments of T, the total number of runs of all colours, follow directly from first principles. Define a characteristic random variable α_t which has the property that it equals unity when the balls on either side of the tth gap are the same colour and zero when they are of different colours. Define S by

$$S = r - T = \sum_{t=1}^{r-1} \alpha_t.$$

The moments of S can now, in principle, be written down. Thus

$$\mathscr{E}(\alpha_t) = \sum_{i=1}^{k} \frac{r_i(r_i-1)}{r(r-1)}$$

and

$$\mathscr{E}(S) = \frac{1}{r} \sum_{i=1}^{k} r_i(r_i-1).$$

We shall write generally

$$F_\omega = \sum_{i=1}^{k} r_i^{(\omega)},$$

so that in this notation

$$r.\mathscr{E}(S) = F_2.$$

Again

$$\mathscr{E}[S(S-1)] = \mathscr{E}\left[\left(\sum_{t=1}^{r-1} \alpha_t\right)^2 - \sum_{t=1}^{r-1} \alpha_t\right].$$

Because α_t is a characteristic random variable

$$\mathscr{E}(\alpha_t^m) = \mathscr{E}(\alpha_t),$$

but it will be necessary to divide the double sum into two parts:

$$\mathscr{E} \sum_{\substack{t=1 \\ t \neq \omega}}^{r-1} \sum_{\omega=1}^{r-1} \alpha_t \alpha_\omega = \mathscr{E} \sum_{t=1}^{r-1} \alpha_t \alpha_{t+1} + \mathscr{E} \sum_{\substack{t=1 \\ |t-\omega| \geqslant 2}}^{r-1} \sum_{\omega=1}^{r-1} \alpha_t \alpha_\omega.$$

The expected values of the products are

$$\mathscr{E}(\alpha_t \alpha_{t+1}) = \sum_{i=1}^{k} \frac{r_i(r_i-1)(r_i-2)}{r(r-1)(r-2)}$$

and

$$\mathscr{E}(\alpha_t \alpha_\omega) = \sum_{i \neq j} \frac{r_i(r_i-1) r_j(r_j-1)}{r(r-1)(r-2)(r-3)} + \sum_{i=1}^{k} \frac{r_i(r_i-1)(r_i-2)(r_i-3)}{r(r-1)(r-2)(r-3)},$$

there being $2(r-2)$ products in which the suffices of the α's differ by unity and $(r-2)(r-3)$ products in which the suffices differ by two or more. From David & Kendall's tables of symmetric functions the double sums in the expectations can be eliminated and we have finally

$$r^{(2)} \mu_{(2)}(S) = F_2(F_2-2) - 2F_3.$$

By a similar process we obtain the third and fourth factorial moments of S which are

$$r^{(3)} \mu_{(3)}(S) = F_2(F_2-2)(F_2-4) - 6F_3(F_2-3) + 10F_4 + 10F_3,$$

$$r^{(4)} \mu_{(4)}(S) = F_2(F_2-2)(F_2-4)(F_2-6) - 12F_3(F_2-3)(F_2-5) + 40F_4(F_2-4) + 40F_3(F_2-3) + 12F_3(F_3-3) - 84F_5 - 296F_4 - 120F_3.$$

5. Two approaches to the limit are possible. First fix k, the number of colours, and let r, the total number of balls, increase without limit. Then we have that

$$\frac{S-F_2/r}{\sigma_S}$$

is in the limit a unit normal variable where

$$\sigma_S^2 = \frac{F_2(r-3)}{r(r-1)} + \frac{F_2^2}{r^2(r-1)} - \frac{2F_3}{r(r-1)}.$$

This result follows by generalizing Mood's procedure for runs of one colour. As an alternative we can fix the possible numbers of each colour, i.e. we can put an upper bound on r_i, and let k and therefore r tend to infinity. Under these conditions we have approximately

$$\mu_{(w)}(S) = \mathscr{E}\Big(\sum_{i_1+i_2+\ldots+i_w} \alpha_{i_1}\alpha_{i_2}\ldots\alpha_{i_w}\Big) \doteq \sum \frac{r_{i_1}^{(i_1+1)}\ldots r_{i_w}^{(i_w+1)}}{r^{\left(\sum_{j=1}^{w} i_j + w\right)}},$$

since the other sums in which some or all of the subscripts are equal will be of lower order in r. Accordingly if we put

$$\lambda = F_2/r,$$

then

$$\mu_{(w)}(S) \to \lambda^w,$$

which is the wth Poisson factorial moment. This indicates that S tends in the limit to be distributed as a Poisson variable with parameter λ. A rigorous proof of this can be given following the lines already set out elsewhere for moments of a similar structure (Barton, 1957). We have therefore that as k (and r) increase without limit the distribution of the total number of runs tends to that of Poisson's binomial limit with parameter F_2/r. In practice this limit will not be reached until k is large.

6. Following Aitken (1939) we define the factorial cumulants $\kappa_{(l)}$ by the relation

$$\kappa_l = \sum_{i=1}^{l} \frac{\Delta^i(0)^l}{i!}\kappa_{(i)},$$

where κ_l is the ordinary moment cumulant. The first four of these are

$$\kappa_1 = \kappa_{(1)}, \quad \kappa_2 = \kappa_{(2)} + \kappa_{(1)}, \quad \kappa_3 = \kappa_{(3)} + 3\kappa_{(2)} + \kappa_{(1)}, \quad \kappa_4 = \kappa_{(4)} + 6\kappa_{(3)} + 7\kappa_{(2)} + \kappa_{(1)}.$$

From these relations we may find the factorial cumulants of S:

$$\kappa_1(S) = \frac{F_2}{r}, \quad \kappa_{(2)} = \frac{F_2^2 - 2rF_2 - 2rF_3}{r^2(r-1)},$$

$$\kappa_{(3)}(S) = \frac{1}{r^3(r-1)(r-2)}[4F_2^3 - 12rF_2^2 + 8r^2F_2 - 12rF_3F_2 + 28r^2F_3 + 10r^2F_4].$$

The expression for $\kappa_{(4)}(S)$ is lengthy and we do not reproduce it here. Write

$$\lambda = F_2/r, \quad \lambda_l = F_{2+l}/r,$$

whence

$$\kappa_1(S) = \lambda, \quad \kappa_{(2)}(S) = \frac{\lambda^2 - 2\lambda - 2\lambda_1}{r-1}, \quad \kappa_{(3)}(S) = \frac{1}{(r-1)(r-2)}[4\lambda^{(3)} - 6\lambda_1(2\lambda - 3) + 10(\lambda_1 + \lambda_2)].$$

These cumulants simplify considerably if we let

$$r_i = R \quad (i = 1, 2, \ldots, k),$$

for in this case

$$\lambda_t = \lambda^{(t+1)}.$$

Thus for equal numbers of each colour we have

$$\kappa_1 = \lambda, \quad \kappa_{(2)} = -\frac{\lambda^2}{r-1}, \quad \kappa_{(3)} = \frac{2\lambda^2(\lambda-1)}{(r-1)(r-2)}, \quad \kappa_{(4)} = -\frac{6\lambda^2}{(r-1)(r-2)(r-3)}\left[\lambda^2 - 4\lambda + 2 + \frac{\lambda^2}{r-1}\right].$$

If we consider a simple binomial expression, say

$$(q+p)^r,$$

where

$$\lambda = rp,$$

the factorial moments are

$$\kappa_1 = \lambda, \quad \kappa_{(2)} = -\frac{\lambda^2}{r}, \quad \kappa_{(3)} = \frac{2\lambda^3}{r^2}, \quad \kappa_{(4)} = -\frac{6\lambda^4}{r^3}.$$

It will be noted that the leading term in the factorial cumulant of S is the same as the factorial cumulants of the binomial in the four cases given.

7. The similarity between the first four cumulants of the distribution of S, when the numbers of each colour are the same, and those of the simple binomial, coupled with the fact that two possible distributions under conditions analogous to those of proceeding to binomial limits are the normal distribution and Poisson, suggests that a binomial distribution may be a useful approximation to the distribution of S. The case $r = 12$ with four (k) colours, each three (R) in number was considered. S can take values 0, 1, 2, ..., 8, and

$$\kappa_1(S) = \frac{F_2}{r} = R - 1, \quad \kappa_2(S) = \frac{F_2^2}{r^2(r-1)} + \frac{F_2(r-3)}{r(r-1)} - \frac{2F_3}{r(r-1)} = \frac{(R-1)(r-R)}{r-1}.$$

Three approximations to the distribution of S are now considered. First we let the binomial index equal 8, and fit the binomial

$$(q+p)^n = (\tfrac{3}{4} + \tfrac{1}{4})^8,$$

the 'p' of the binomial being found from

$$8p = \kappa_1 = 2.$$

This is approximation I of Table 3. Secondly, we equate the first two moments of S, i.e. $R-1$ and $(R-1)(r-R)/(r-1)$, to the first two binomial moments, which is equivalent to taking

$$p = \frac{R-1}{r-1} = \frac{2}{11}, \quad n = \frac{\kappa_1}{p} = r - 1 = 11.$$

This is approximation II. Thirdly, we may calculate areas corresponding to a normal curve with the correct mean and variance of S. This is approximation III. The approximate and exact distributions are given in Table 3, the normal curve area in the first group representing the whole left-hand tail. Approximation II—the simple binomial with the correct first two moments of S—is clearly the best, but no mistake is likely in a test of significance whichever approximation is used.

7. Although it is possible to show similarity between the factorial binomial cumulants and those of the distribution of S in a simple way when there are equal numbers of each colour, the approximations to the distribution of S are so good for this particular case that one might expect them to be reasonable for the case when the numbers of each colour are not the same. We found that this was so. It is supposed there are 6 balls of one colour, 4 of another and 1 of each of two other colours. Thus if we consider approximation I and assume

$$n = 8, \quad \kappa_1 = \frac{1}{r}\sum_i r_i(r_i - 1) = 3{\cdot}5, \quad p = \tfrac{7}{16},$$

and the binomial is $(\frac{9}{16} + \frac{7}{16})^8$.

Table 3. *Comparison of true distribution of S with three approximations* ($r = 12$. [3^4])

S	0	1	2	3	4	5	6	7	8
Approximation I	0·1001	0·2670	0·3115$^-$	0·2076	0·0865$^+$	0·0231	0·0038	0·0004	0·0000
Approximation II	0·1100	0·2689	0·2987	0·1992	0·0885$^+$	0·0275$^+$	0·0061	0·0010	0·0001
Approximation III	0·1205	0·2274	0·3042	0·2274	0·0952	0·0222	0·0029	0·0002	—
True	0·1118	0·2670	0·2966	0·2003	0·0903	0·0275$^+$	0·0057	0·0008	0·0001

Table 4. *Comparison of true distribution of S with three approximations* ($r = 12$. [641^2])

S	0	1	2	3	4	5	6	7	8
Approximation I	0·0100	0·0624	0·1698	0·2641	0·2567	0·1598	0·0621	0·0138	0·0013
Approximation II	0·0083	0·0563	0·1654	0·2714	0·2697	0·1632	0·0564	0·0091	0·0002
Approximation III	0·0126	0·0552	0·1599	0·2723	0·2723	0·1599	0·0552	0·0112	0·0013
True	0·0054	0·0574	0·1688	0·2749	0·2673	0·1582	0·0567	0·0104	0·0009

For approximation II

$$\kappa_2 = \frac{(r-3)}{r(r-1)}\Sigma r_i(r_i-1) + \left(\frac{r-3}{r(r-1)}\right)^2 (\Sigma r_i(r_i-1))^2 - \frac{2}{r(r-1)}\Sigma r_i(r_i-1)(r_i-2)$$

$$= \tfrac{79}{44},$$

giving the binomial

$$(\tfrac{79}{154} + \tfrac{75}{154})^{539/75} = (0{\cdot}51299 + 0{\cdot}48701)^{7{\cdot}186}.$$

The fractional value of the index was used to calculate the probabilities. In the third approximation it was assumed that S is normally distributed with mean 3·5 and variance 79/44, the tail areas being put together in each of the two end groups. The agreement is surprisingly good considering that r is small and that there is disparity between the numbers of each colour. The results would suggest that for $r > 12$, whatever the composition of the numbers of colours, the normal approximation (III) will be adequate for tests of significance.

8. An extension of the theory of runs can be used in what we might call a test for the persistence of type. Assuming the r events of k possible types, the null hypothesis will be that these r events are a sample from a multinomial, the probability of the ith type of which is p_i. This is to say that under H_0 we assume

$$p(r_1, r_2, \ldots, r_k) = \frac{r!}{\prod_{i=1}^{k} r_i!} \prod_{i=1}^{k} p_i^{r_i},$$

with each of the

$$\frac{r!}{\prod_{i=1}^{k} r_i!}$$

sequences being equally likely. The alternate hypothesis, H_1, might be that we have a simple Markoff chain in which the probability of getting an event of the ith type at any given drawing is greater if the immediately preceding drawing also is of the ith type and less if otherwise. We may write these transition probabilities

$$p_{ii} = p_i^2(1+\theta),$$

$$p_{ij} = p_i p_j(1-\theta W) \quad (i \neq j),$$

where

$$W = \Sigma p_i^2/(1-\Sigma p_i^2).$$

Assuming $\theta = 0$ for the first drawing of the sample, the probability of any given sequence in which there are T groups is

$$(1+\theta)^{r-T-1}(1-\theta W)^T \prod_{i=1}^{k} p_i^{r_i}.$$

If we now consider the conditional distribution under H_1 for a sequence of given composition of numbers of each type, this is equal to

$$\left(\frac{1-\theta W}{1+\theta}\right)^T K,$$

where K is a factor of proportionality and depends only on $r_1, \ldots, r_k$ and

$$\frac{1-\theta W}{1+\theta} = \phi \quad \text{(say)}.$$

It follows that T is sufficient for ϕ and that the likelihood ratio test for $\theta = 0$, or equivalently $\phi = 0$, is a function of T. Thus the use of T is equivalent to the likelihood ratio test in this case.

9. We have been interested in the multiple runs problem for its own sake but, apart from the simple stochastic problem of the preceding section, it is perhaps worth while to describe one possible analysis of variance application. The use of runs in statistical theory is obvious, but there appears room for them in the category which Tukey has aptly described elsewhere as 'quick and dirty'. By quick and dirty tests we would understand tests which are easy to apply but probably of low power, so that any effect which is significant using them will certainly be more significant with a more sensitive test. On the other hand, if the looked-for effect is not significant with the 'quick and dirty' test this does not necessarily mean that it cannot be picked out by more refined methods. We illustrate the working on an example of

the breaking strength of cement-mortar briquettes (Table 5). We mark the observations in order of magnitude, ties being decided by tossing a coin. The ranking is simply in order to enable the runs to be counted more easily. Each group will correspond to a different colour. The number of runs is 22 and

$$S = 25 - 22 = 3, \quad \mathscr{E}(S) = 4, \quad \text{var}(S) = \tfrac{10}{3}.$$

Assuming S to be normally distributed we see that significance is not achieved, a result which in this case, is confirmed by the orthodox F-test.

Table 5. *Breaking strength of cement-mortar briquettes*

Group	1	2	3	4	5
Tension in lb.	518 (5)	508 (3)	538 (12)	535 (9)	492 (1
	560 (20)	574 (23)	544 (15)	540 (14)	506 (2
	538 (11)	528 (6)	554 (18)	550 (17)	528 (7
	510 (4)	534 (8)	579 (24)	555 (19)	536 (10
	544 (6)	538 (13)	598 (25)	567 (21)	572 (2

10. Although the analysis of variance runs test has little power this will not be the case for all applications. We owe the following elegant illustration to E. S. Pearson. Consider the falls in the price of shares on the London Stock Exchange during the period 6 November to 8 December 1956, both dates inclusive. Five types of industrial activity, *A* Insurance, *B* Breweries and Distilleries, *C* Electrical Equipment and Radio, *D* Motor and Aircraft, *E* Oil were chosen for study. The closing prices as given in *The Times* for eighteen businesses of each type were taken, and Table 6 shows, for each day, the type of industrial activity for which the greatest number of the eighteen showed a fall in price from the previous day. In the few cases where there were equal numbers for two types, that type which also showed the fewer rises in price was taken:

Table 6. *Type of industrial activity showing greatest number of falls in price of shares*

6 Nov. *A*	13 Nov. *B*	20 Nov. *E*	27 Nov. *B*	4 Dec. *C*
7 Nov. *A*	14 Nov. *C*	21 Nov. *C*	28 Nov. *E*	5 Dec. *C*
8 Nov. *D*	15 Nov. *C*	22 Nov. *E*	29 Nov. *A*	6 Dec. *D*
9 Nov. *D*	16 Nov. *C*	23 Nov. *E*	30 Nov. *E*	7 Dec. *C*
10 Nov. *A*	17 Nov. *E*	24 Nov. *E*	1 Dec. *E*	8 Dec. *B*

We have $r = 25$ with $r_A = 4$, $r_B = r_D = 3$, $r_C = 7$, $r_E = 8$, $T = 16$ and $S = r - T = 9$. From formulae we find

$$F_2 = \sum_{i=1}^{5} r_i(r_i - 1) = 122, \quad F_3 = \sum_{i=1}^{5} r_i(r_i - 1)(r_i - 2) = 582,$$

$$\mu_2(S) = \frac{F_2(r-3)}{r(r-1)} + \frac{F_2^2}{r^2(r-1)} - \frac{2F_3}{r(r-1)} = 1{\cdot}54106,$$

$$\sigma(S) = 1{\cdot}241 \quad \text{and} \quad \mathscr{E}(S) = 4{\cdot}88,$$

so that

$$\frac{S - \mathscr{E}(S)}{\sigma_S} = \frac{4{\cdot}12}{1{\cdot}241} > 3.$$

This normal deviate is significant, and we conclude therefore that the test is picking out the fact that during the Middle East crisis there was some persistence from day to day in the way in which different classes of shares were affected.

11. In nearly all statistical applications of the theory of runs the test of significance will be one-tailed. For example, in Wald & Wolfowitz's application of the two-colour distribution to the two-sample means test, a small number of runs could be held to indicate a possible separation of the population means. The critical region in their case will be the tail where T is small or S large. This is also true for the applications given above; for instance, in the analysis of variance application a complete separation of the twenty-five observations into five groups would indicate the possibility of the five population means being different instead of equal under the null hypothesis. We choose therefore the lower tail of T or the upper tail of S. The result when there is too great an alternation of colour and T is significant at the upper tail has not so obvious a statistical interpretation.

REFERENCES

Aitken, A. C. (1939). *Statistical Mathematics.* Edinburgh: Oliver & Boyd.
Barton, D. E. (1957). *J. R. Statist. Soc.* B (to appear).
Mood, A. M. (1940). *Ann. Math. Statist.* **11**, 367.
Stevens, W. L. (1939). *Ann. Eugen., Lond.*, **9**, 10.
Wald, A. & Wolfowitz, J. (1940). *Ann. Math. Statist.* **11**, 112.
Whitworth, W. A. (1886). *Choice and Chance.* Cambridge: Deighton Bell and Co.

Table 1. *Three-colour runs* ($S = r - T$)

(The probabilities are obtained by dividing the number tabled by the corresponding value of $\frac{r!}{\Pi r_i!}$)

r	r_1	r_2	r_3	$\frac{r!}{\Pi r_i!}$	Values of T									
					3	4	5	6	7	8	9	10	11	12
6	2	2	2	90	6	18	36	30						
	3	2	1	60	6	18	26	10						
	4	1	1	30	6	18	6	—						
7	3	2	2	210	6	24	62	80	38					
	3	3	1	140	6	24	52	40	18					
	4	2	1	105	6	24	42	30	3					
	5	1	1	42	6	24	12	—	—					
8	3	3	2	560	6	30	100	180	170	74				
	4	2	2	420	6	30	90	150	120	24				
	4	3	1	280	6	30	80	90	60	14				
	5	2	1	168	6	30	60	60	12	—				
	6	1	1	56	6	30	20	—	—	—				
9	3	3	3	1680	6	36	150	360	510	444	174			
	4	3	2	1260	6	36	140	310	405	284	79			
	5	2	2	756	6	36	120	240	252	96	6			
	4	4	1	630	6	36	120	180	180	84	24			
	5	3	1	504	6	36	110	160	132	56	4			
	6	2	1	252	6	36	80	100	30	—	—			
	7	1	1	72	6	36	30	—	—	—	—			
10	4	3	3	4200	6	42	202	580	1050	1234	838	248		
	4	4	2	3150	6	42	192	510	870	894	498	138		
	5	3	2	2520	6	42	182	470	752	692	332	44		
	5	4	1	1260	6	42	162	300	372	252	108	18		
	6	2	2	1260	6	42	152	350	440	240	30	—		
	6	3	1	840	6	42	142	250	240	140	20	—		
	7	2	1	360	6	42	102	150	60	—	—	—		
	8	1	1	90	6	42	42	—	—	—	—	—		
11	4	4	3	11550	6	48	266	900	2010	3064	3012	1764	480	
	5	3	3	9240	6	48	256	840	1802	2568	2340	1168	212	
	5	4	2	6930	6	48	246	750	1527	1968	1548	702	135	
	6	3	2	4620	6	48	226	660	1220	1360	870	220	10	
	5	5	1	2772	6	48	216	480	744	672	432	144	30	
	6	4	1	2310	6	48	206	450	645	560	300	90	5	
	7	2	2	1980	6	48	186	480	690	480	90	—	—	
	7	3	1	1320	6	48	176	360	390	280	60	—	—	
	8	2	1	495	6	48	126	210	105	—	—	—	—	
	9	1	1	110	6	48	56	—	—	—	—	—	—	
12	4	4	4	34650	6	54	342	1350	3618	6894	9036	7938	4320	1092
	5	4	3	27720	6	54	332	1270	3300	5974	7388	5982	2826	588
	6	3	3	18480	6	54	312	1140	2778	4570	5060	3360	1100	100
	5	5	2	16632	6	54	312	1080	2592	4104	4272	2880	1110	222
	6	4	2	13860	6	54	302	1030	2388	3620	3550	2130	710	70
	7	3	2	7920	6	54	272	880	1818	2350	1820	660	60	—
	6	5	1	5544	6	54	272	700	1320	1400	1120	540	110	22
	7	4	1	3960	6	54	252	630	1008	1050	660	270	30	—
	8	2	2	2970	6	54	222	630	1008	840	210	—	—	—
	8	3	1	1980	6	54	212	490	588	490	140	—	—	—
	9	2	1	660	6	54	152	280	168	—	—	—	—	—
	10	1	1	132	6	54	72	—	—	—	—	—	—	—

Table 2. *Four-colour runs* ($S = r - T$)

(The probabilities are obtained by dividing the number tabled by the corresponding value of $\frac{r!}{\Pi r_i!}$)

r	r_1	r_2	r_3	r_4	$\frac{r!}{\Pi r_i!}$	Values of T								
						4	5	6	7	8	9	10	11	12
6	2	2	1	1	180	24	72	84						
	3	1	1	1	120	24	72	24						
7	2	2	2	1	630	24	108	252	246					
	3	2	1	1	420	24	108	192	96					
	4	1	1	1	210	24	108	72	6					
8	2	2	2	2	2520	24	144	504	984	864				
	3	2	2	1	1680	24	144	444	684	384				
	3	3	1	1	1120	24	144	384	384	184				
	4	2	1	1	840	24	144	324	294	54				
9	3	2	2	2	7560	24	180	780	2010	2880	1686			
	3	3	2	1	5040	24	180	720	1560	1720	836			
	4	2	2	1	3780	24	180	660	1320	1260	336			
	4	3	1	1	2520	24	180	600	870	660	186			
	5	2	1	1	1512	24	180	480	600	216	12			
10	3	3	2	2	25200	24	216	1140	3720	7480	8416	4204		
	3	3	3	1	16800	24	216	1080	3120	5160	5016	2184		
	4	2	2	2	18900	24	216	1080	3330	6210	6066	1974		
	4	3	2	1	12600	24	216	1020	2730	4170	3366	1074		
	5	2	2	1	7560	24	216	900	2160	2736	1368	156		
	4	4	1	1	6300	24	216	900	1740	1980	1116	324		
	5	3	1	1	5040	24	216	840	1560	1536	768	96		
11	3	3	3	2	92400	24	252	1584	6360	16680	27756	27408	12336	
	4	3	2	2	69300	24	252	1524	5820	14400	22056	18708	6516	
	4	3	3	1	46200	24	252	1464	5070	10720	14256	10848	3566	
	5	2	2	2	41580	24	252	1404	4950	11016	14184	8364	1386	
	4	4	2	1	34650	24	252	1404	4350	9000	10656	6768	2016	
	5	3	2	1	27720	24	252	1344	4200	7896	8484	4704	816	
	5	4	1	1	13860	24	252	1224	2910	4176	3384	1584	306	
	6	2	2	1	13860	24	252	1164	3210	4920	3480	780	30	
	6	3	1	1	9240	24	252	1104	2460	2920	1980	480	20	
12	3	3	3	3	369600	24	288	2112	10176	33360	74016	109632	98688	41304
	4	3	3	2	277200	24	288	2052	9486	29590	61916	83952	66638	23254
	4	4	2	2	207900	24	288	1992	8796	26100	51216	62892	43128	13464
	5	3	2	2	166320	24	288	1932	8316	23856	44520	50484	30468	6432
	4	4	3	1	138600	24	288	1932	7896	20580	35616	39312	25428	7524
	5	3	3	1	110880	24	288	1872	7416	18616	30320	31104	17528	3712
	5	4	2	1	83160	24	288	1812	6726	15966	23820	21384	10818	2322
	6	2	2	2	83160	24	288	1752	6876	17460	27120	22080	7020	540
	6	3	2	1	55440	24	288	1692	5976	13060	17120	12780	4160	340
	5	5	1	1	33264	24	288	1632	4656	8352	9024	6336	2448	504
	6	4	1	1	27720	24	288	1572	4386	7410	7620	4680	1590	150

BINOMIAL SAMPLING SCHEMES AND THE CONCEPT OF INFORMATION

By D. V. LINDLEY

Statistical Laboratory, University of Cambridge

1. *Summary.* Methods of sampling a binomial population in order to obtain a prescribed accuracy in the determination of the unknown proportion are discussed. The concept of information due to Shannon is used, and the relationship between it and the corresponding concept of Fisher's is investigated and used to explain certain features of the sampling schemes.

2. We consider random sampling from a population in which each individual can be classed as A or not-A, the unknown proportion of A's in the population being θ. The problem discussed is how much sampling should be done in order to obtain sufficient knowledge, in some sense, about the value of θ. In a previous paper (Lindley, 1956) we have proposed a solution using the idea of information, and in the present paper we discuss in detail the application of this idea to binomial sampling.

It is supposed that at any stage of the sampling the knowledge of θ can be expressed by a probability distribution $p(\theta)$. If this is so the quantity

$$I_\theta = I(p(\theta)) = \int p(\theta) \log p(\theta)\, d\theta \tag{1}$$

provides a convenient measure of the *amount of information* about θ; indeed, accepting a few mild requirements on the properties that 'amount of information' should possess, it has been shown by Shannon that I_θ is, apart from an arbitrary multiplying constant, unique. (The constant is incorporated in the arbitrariness of the base of the logarithm.) Details will be found in the paper already cited. The effect of sampling will be to alter $p(\theta)$, according to Bayes's theorem, and so to change I_θ. The rule we propose is: *continue sampling until I_θ has reached some prescribed value.*

To facilitate the calculations it will be supposed that $p(\theta)$ belongs to the family

$$p_{ab}(\theta) = \theta^{a-1}(1-\theta)^{b-1}\,\Gamma(a+b)/\{\Gamma(a)\,\Gamma(b)\}, \tag{2}$$

with a and b positive. This family of densities has the property that if the prior distribution is $p_{ab}(\theta)$ then the posterior distribution, after a single binomial trial has been performed, is $p_{a+1,b}(\theta)$ or $p_{a,b+1}(\theta)$ according as the result of the trial was A or not-A respectively. This fact permits representation of the sampling schemes on a diagram with axes giving the values of a and b. Simple calculations show that

$$\begin{aligned} I(p_{ab}(\theta)) &= \ln[\Gamma(a+b)/\{\Gamma(a)\,\Gamma(b)\}] \\ &\quad + (a-1)[\Psi(a)-\Psi(a+b)] + (b-1)[\Psi(b)-\Psi(a+b)], \end{aligned} \tag{3}$$

where natural logarithms have been used in (1) and

$$\Psi(x) = d \ln \Gamma(x)/dx.$$

Application of the sampling rule proposed involves the sampling continuing until the values of a and b obtained are such that (3) has attained a prescribed value. We defer

consideration of the details of this for a moment in order to note that the integral (1) is not invariant under a change of description of the parameter values. Specifically if $\phi = \phi(\theta)$ is a monotone function of θ then the probability distribution, $q(\phi)$, of ϕ will be given by

$$q(\phi)\,d\phi = p(\theta)\,d\theta, \tag{4}$$

and

$$\begin{aligned} I_\phi = I(q(\phi)) &= \int q(\phi)\log q(\phi)\,d\phi \\ &= \int p(\theta)\log\left\{p(\theta)\frac{d\theta}{d\phi}\right\}d\theta \\ &= I_\theta + \int p(\theta)\log\frac{d\theta}{d\phi}\,d\theta. \end{aligned} \tag{5}$$

It follows that if a sampling scheme is adopted in which we try to obtain a prescribed amount of information about ϕ, it will, in general, be different from a scheme relevant to θ. We consider three schemes depending on the choice of ϕ.

3. *Information about θ.* The information about θ is given by (3). This expression, as it stands, is too complicated to be easily understood but considerable simplification is possible by use of the standard asymptotic formulae

$$\ln\Gamma(x) \sim \ln\sqrt{(2\pi)} - x + (x-\tfrac{1}{2})\ln x,$$

Stirling's formula, and

$$\Psi(x) \sim \ln x - 1/(2x).$$

We obtain

$$I(p_{ab}(\theta)) \sim \tfrac{1}{2}\ln\{(a+b)^3/(2\pi ab)\} - \tfrac{1}{2},$$

for large values of both a and b. It is, however, worth noting that both the asymptotic formulae are remarkably accurate, and that therefore when we refer to large a and b we often only mean that they are both greater than 5. It follows that the boundary in the (a, b) diagram is a curve of the form

$$(a+b)^3 = \lambda ab, \tag{6}$$

where λ is a constant dependent on the amount of information required. Thus if the prior distribution is such that the point (a, b) lies to the 'south-west' of the curve in the first quadrant, sampling is continued until the curve is crossed. The general features of the boundary (drawn roughly in Fig. 1) and its form for small a and b have been discussed previously (Lindley, 1956). The main point of interest is the manner in which the boundary approaches the origin when either a or b is small; the approach is faster than even (6) would suggest. This agrees with the 'common-sense' feeling that if the same thing continually happens, say the sun rises each morning, then we are much better informed than we would be if there was known to be even a single non-occurrence.

4. *Information about $\phi = 2\arcsin\sqrt{\theta}$.* Here $d\theta/d\phi = \{\theta(1-\theta)\}^{\frac{1}{2}}$, and on evaluating the final term in (5) and adding it to I_θ given by (3) we obtain

$$I_\phi = \ln[\Gamma(a+b)/\{\Gamma(a)\,\Gamma(b)\}] + (a-\tfrac{1}{2})[\Psi(a) - \Psi(a+b)] + (b-\tfrac{1}{2})[\Psi(b) - \Psi(a+b)]. \tag{7}$$

By use of the asymptotic formulae it can be shown that for large a and b

$$I_\phi \sim \tfrac{1}{2}\ln\{(a+b)/(2\pi)\} - \tfrac{1}{2}, \tag{8}$$

and hence the boundary of the region required is $a+b =$ constant, that is a fixed sample size scheme. When, however, either a or b are small the asymptotic result does not provide

a good approximation. Consider the situation when $a = 1$. If b is large the asymptotic expressions yield

$$I_\phi \sim \tfrac{1}{2}\ln b - \tfrac{1}{2}\gamma - 1 \tag{9}$$

instead of (8). (γ is Euler's constant, equal to $\Psi(1)$.) If it has been decided to use a prescribed amount of information resulting, when a and b are large, in a fixed sample of size n, the critical amount of information when $a = 1$ will be attained by a value of b satisfying

$$\tfrac{1}{2}\ln b - \tfrac{1}{2}\gamma - 1 = \tfrac{1}{2}\ln\frac{n}{2\pi} - \tfrac{1}{2},$$

on equating (8) and (9). Hence

$$b = n[e^{1+\gamma}/(2\pi)]. \tag{10}$$

The constant in square brackets is about 0·77. In other words if $a = 1$ only about three-quarters the amount of sampling (every observation being not-A) is necessary to obtain the same amount of information about 2 arc sin $\sqrt{\theta}$ as when a is large. Thus, for example, if we are initially ignorant about θ we should presumably take $a = b = 1$: or if ignorant about ϕ, $a = b = \frac{1}{2}$, it makes little difference which. In either case we should obtain the requisite information about ϕ more quickly if every trial resulted in not-A than we would if A's and not-A's were mixed. This is the same phenomenon that was observed when acquiring information about θ but here it is not nearly so striking. The constant size boundary is adequate for values of a (or symmetrically of b) greater than 5. Thus when $a = 5$ the constant factor corresponding to that in (10) is 0·96. The general form of the boundary is shown in Fig. 2.

5. *Information about* $\psi = \ln\{\theta/(1-\theta)\}$. The consideration of a constant amount of information for 2 arc sin $\sqrt{\theta}$ instead of θ is equivalent to asking for more information about θ when it is near 0 or 1 than when it is near $\frac{1}{2}$. Such a tendency can often be exaggerated still further with profit. Suppose we imagine the continuous range of values of θ ($0 \leqslant \theta \leqslant 1$) replaced by a finite set of values so chosen that it would be important to distinguish between two neighbouring values of θ but finer distinctions would be unimportant. Then if $\Delta\theta$ denotes this minimal distinction it will often happen that $\Delta\theta/\theta$ would be constant for θ near 0; that is the percentage accuracy is fixed. Over the whole range of θ, $\Delta\theta/\{\theta(1-\theta)\}$ might be constant, a form which reduces to the previous for small θ and which is symmetrical in A and not-A. Let us find a transform, ψ, of θ, such that the same distinctions in ψ are required over its entire range; then

$$\Delta\psi = \frac{\Delta\theta}{\theta(1-\theta)},$$

or, in passing to the limit,

$$\psi = \ln\{\theta/(1-\theta)\}. \tag{11}$$

The same transformation arises naturally in another way, though its connexion with the present study is not clear and may be accidental. It is well known that the only distributions admitting a single minimal sufficient statistic for all sample sizes are those belonging to the exponential family, that is with likelihood of the form

$$H(x)\,G(\psi)\,e^{u\psi},$$

where u is the minimal sufficient statistic, a function of the sample values x, and ψ is the parameter. The binomial distribution belongs to this family with $\psi = \ln\{\theta/(1-\theta)\}$. It is worth noting that for any number of repetitions of a general binomial sample scheme the

statistics Σa_i and Σb_i are jointly sufficient. In most cases they will also be minimal sufficient but in exceptional cases this will not happen. Thus in the fixed sample-size scheme and the Haldane inverse binomial scheme (see below) Σb_i is alone minimal sufficient. This is because Σa_i can, in those two special cases, be deduced from the sizes of the schemes (which are not statistics) and, in the former case, Σb_i.

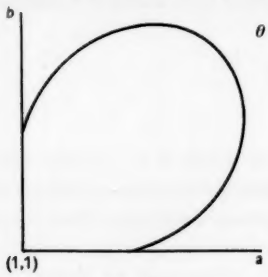


Fig. 1

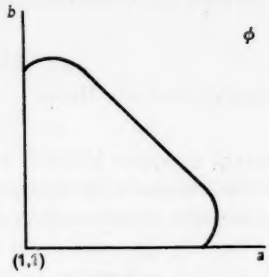


Fig. 2

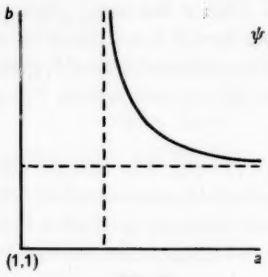


Fig. 3

Figs. 1–3. Sampling schemes for providing a prescribed amount of information about the parameters indicated. For the notation see text.

Considering, then, schemes which produce constant information for ψ. We have $d\theta/d\psi = \theta(1-\theta)$ and on evaluating the final term in (5) and adding it to I_θ given by (3) we obtain

$$I_\psi = \ln[\Gamma(a+b)/\{\Gamma(a)\,\Gamma(b)\}] + a[\Psi(a) - \Psi(a+b)] + b[\Psi(b) - \Psi(a+b)]. \tag{12}$$

The asymptotic results used before yield

$$I_\psi \sim \tfrac{1}{2}\ln[ab/\{2\pi(a+b)\}] - \tfrac{1}{2}, \tag{13}$$

and hence the boundary of the region required is

$$\frac{ab}{a+b} = N, \quad \text{a constant}, \tag{14}$$

a rectangular hyperbola with axes $a = N$, $b = N$: only the portion in the quadrant $a \geqslant N$, $b \geqslant N$ will be meaningful in the present situation. This obtains provided a and b are both greater than about 5. Now, from (12), we have

$$\partial I_\psi/\partial a = a\Psi'(a) - (a+b)\,\Psi'(a+b),$$

which is positive since $\frac{d}{dx}[x\Psi'(x)] < 0$. Hence the information about ψ always increases when an additional observation is incorporated; with the two previous parameters we saw that this did not necessarily happen for small a and b. Thus the phenomenon connected with constant repetition does not obtain here. From this it follows that (14) is a good approximation to the boundary provided $N \geqslant 5$, that is provided the required amount of information is sufficiently large. The general form is shown in Fig. 3. The case $N < 5$ does not seem of sufficient importance to warrant a numerical investigation.

In situations where it is appropriate to consider constancy of information about ψ, we would often have prior knowledge to suggest that θ is small, and it would probably be convenient to express this prior knowledge of θ by a distribution belonging to the family (2) with a small and b large. If it then happened that $a < N$ and $b > N$ the boundary (14) might conveniently be replaced by the asymptote to the hyperbola, namely $a = N$, without seriously affecting the approximation. This modified sampling scheme is that advocated by Haldane (1945) in just the circumstances we have been considering. He was concerned with producing a sampling scheme which would yield an approximately constant coefficient of variation for the estimate of θ. It is easy to see that for small θ this is equivalent to requiring the standard deviation of ψ to be constant when ψ is large. That this will be so in large samples follows immediately from the considerations in the next section.

It is possible to consider the transformation $\tau = \ln \theta$, constant information about which more closely resembles the situation considered by Haldane. This causes trouble for small a and b but for large values the boundary is again of the form $a =$ constant.

The three sampling schemes described here are quite different in character, as may most easily be seen by a glance at the three figures. The first and third represent the extreme cases: in the former one is concerned to know θ, whatever its value, to a fixed accuracy, for example to two decimal places; in the latter the proportional accuracy is constant. The second, giving approximately the fixed sample-size scheme, represents an intermediate requirement. The choice of which to use depends essentially on the type of knowledge about θ that is needed. It is interesting that no loss function is involved in the approach, yet this choice involves considerations which are similar in form to those which would be advanced in deciding on a loss function. The reason is that Shannon's integral (1) expresses uniquely the idea of information, the only choice left, so to speak, is the $d\theta$ in the integral. On the other hand the loss function approach is distinct from one based on the integral (1), as has been demonstrated previously (Lindley, 1956). The reader can easily construct other schemes for special situations, using the method given here.

6. *The concepts of information.* The sampling rules introduced here have been constructed using the concept of information due to Shannon. Similar schemes that have previously been suggested have utilized the sampling variance of an efficient estimator as the criterion whereby we can judge whether θ has been found to sufficient accuracy. Essentially, then, the concept of information used has been Fisher's, whose expression

$$\mathscr{I}_\theta = \mathscr{E}_x\left[-\frac{\partial^2 L(x \mid \theta)}{\partial \theta^2}\right], \tag{15}$$

is asymptotically equal to the inverse of the sampling variance referred to. ($L(x \mid \theta)$ denotes the logarithm of the likelihood of the sample point x and $\mathscr{E}_x$ denotes the expectation taken with respect to x.) Now these two concepts of information are totally different in character.

There are two main differences. First, Shannon's concept involves the introduction of a prior distribution for the unknown parameter whereas Fisher's does not. Secondly, Shannon's concept, when applied to construct a sampling scheme, uses the likelihood of the sample point x obtained (in Bayes's theorem), whereas Fisher's concept uses the whole probability distribution of x for fixed θ; that is it involves not merely the x that was obtained but also the x's that might have been obtained. The first point is to the advantage of Fisher's notion, but once the idea of a prior probability for θ has been admitted the second point makes Shannon's concept simpler to apply.

Now despite these great differences we notice that the resulting sampling schemes are closely similar. Thus it is well known that for samples of a fixed size the amount of Fisherian information about ϕ is constant. In other words we can acquire a prescribed amount of Fisherian information about ϕ using a fixed sample-size scheme. But we have just seen that, at least when a and b are not too small, the same result obtains with Shannon's I_θ for $\mathscr{I}_\theta$. Again the Haldane scheme appears in just the same circumstances as it did originally when $\mathscr{I}_\theta$ was used. We now offer an explanation of why this is so.

Bayes's theorem says that

$$p(\theta \mid x) = p(x \mid \theta)\, p(\theta)/p(x),$$

and hence, on taking logarithms,

$$L(\theta \mid x) = L(x \mid \theta) + L(\theta) - L(x).$$

Differentiation twice with respect to θ gives

$$-\frac{\partial^2 L(\theta \mid x)}{\partial \theta^2} = -\frac{\partial^2 L(x \mid \theta)}{\partial \theta^2} - \frac{\partial^2 L(\theta)}{\partial \theta^2}. \tag{16}$$

Suppose that $p(\theta \mid x)$ is approximately a normal distribution, as it is in our study provided a and b are both large, then $I(p(\theta \mid x))$ is easily verified to be approximately $-\frac{1}{2}\ln(2\pi e\sigma^2)$, where σ^2 is the posterior variance of θ, given x, and will, of course, depend on x. It should be distinguished from the variance of an efficient estimator of θ. But, again, in view of the approximate normality, $\sigma^{-2} = -\partial^2 L(\theta \mid x)/\partial\theta^2$, and hence if we choose a sampling scheme, or equivalently a set of points x, such that $I(p(\theta \mid x))$ is constant it will follow that the left-hand side of (16) is constant. For such a sampling scheme if we take expectations over the sample space we shall have, from (16),

$$-\frac{\partial^2 L(\theta \mid x)}{\partial \theta^2} = \mathscr{E}_x\left(-\frac{\partial^2 L(x \mid \theta)}{\partial \theta^2}\right) - \frac{\partial^2 L(\theta)}{\partial \theta^2}.$$

The second term on the right-hand side is negligible in large samples and hence

$$-\frac{\partial^2 L(\theta \mid x)}{\partial \theta^2} \doteq \mathscr{I}_\theta, \tag{17}$$

where $\mathscr{I}_\theta$ denotes Fisher's information which is therefore constant. The same argument persists with ϕ for θ. This demonstrates the equivalence of the two concepts when applied to schemes yielding constant information.

This result establishes the asymptotic equivalence of the posterior variance of θ and the variance of an asymptotically efficient estimator of θ. It is therefore possible to construct the schemes using this latter variance with the asymptotically efficient estimate replacing

the unknown value of θ. Thus the variance of an asymptotically efficient estimator of θ itself is $\theta(1-\theta)/(a+b)$; replacing θ by $a/(a+b)$ yields $ab/(a+b)^3$, which if held constant produces the first of our schemes.

On the other hand in small samples $\mathscr{I}_\theta$ is a difficult concept to handle. Primarily this is because the whole sample space enters into its evaluation: in the binomial situation it is necessary to conduct a tedious enumeration of sample paths. But even when this has been done the solution of the equation $\mathscr{I}_\theta$ = constant presents considerable difficulties. The statistician who refuses to recognize and use prior distributions may still find our method attractive because of this virtue of approximating to another approach. We conclude the paper by discussing the roles of prior distributions and Shannon's concept in the practical application of the schemes.

7. We first note that the prior distributions are not very obtrusive in the present study. They do not effect the boundary of the sampling scheme, but only the point in the (a, b) diagram from which sampling should start. In all practical applications the present author has met the experimenter has a fair idea of the value of θ and this knowledge can usually be expressed approximately by a member of the family of distributions given by (2). For example if he says θ will fairly certainly lie between 10 % and 20 % then a and b can be found such that $p(\theta)$ is only effective over this range. (If fairly certain is interpreted as 95 % certain, then calculations suggest $a = 30$, $b = 175$.) Even if there is considerable prior knowledge, as in this case, experimenters often wish to be able to answer the question: 'What has *this* experiment to say about the value of θ?' They wish to make a statement which does not depend on prior knowledge. Thus Haldane's scheme was devised for cases where θ is known to be small, yet the standard deviation is used to summarize the results and the prior knowledge ignored. In other words, prior considerations enter into the design but not the analysis. It is doubtful if the above question can be answered satisfactorily within a logically consistent framework. Fisherian methods appear to provide an answer but their logical basis is not clear and it is possible to produce contradictions (Mauldon, 1955; Lindley, 1957). It would be possible to provide an answer if a meaning could be given to the phrase 'ignorant about θ'. Although numerous attempts have been made to do this, utilizing the ideas of invariance, none have been conspicuously successful and again contradictions appear (Stein, 1955). Possibly ignorance about θ is as far distant in the past as certainty about θ is into the future. In the three situations discussed in the present paper it does not seem too unreasonable to interpret the phrase 'ignorant about the parameter' to mean that the parameter has a uniform distribution over its permissible range of values. Such an interpretation is equivalent to using the following prior distributions: for θ, $a = b = 1$; for ϕ, $a = b = \frac{1}{2}$; for ψ, $a = b = 0$. (The last is an improper distribution of the type considered by Jeffreys (1948).) It will be noted that even between the two extreme cases (θ and ψ) there is only a difference of effectively two observations: for observations A and not-A would suffice to alter the prior distribution from $a = b = 0$ to $a = b = 1$. The equivalent of two observations is hardly likely to be serious in practical problems. Consequently it may be possible to meet the experimenter's requirement by choosing a and b somewhere around these small values even if the prior knowledge he really possesses suggests otherwise.

Another practical consideration involves the determination of the constant representing the amount of information required, for example the value of N in (14). This is easily done

in the present context. For situations likely to be met in practice the relationship between the information I_θ and the posterior variance of θ, $\sigma^2(\theta)$,

$$I_\theta = -\tfrac{1}{2}\ln\{2\pi e\sigma^2(\theta)\} \tag{18}$$

can be used to evaluate the constant in terms of $\sigma^2(\theta)$, a more familiar concept. For example, if ψ is being considered, we have from (13) and (18)

$$\tfrac{1}{2}\ln\{2\pi e(a+b)/(ab)\} = \tfrac{1}{2}\ln\{2\pi e\sigma^2(\psi)\}$$

and hence $\sigma^2(\psi) = (a+b)/(ab) = 1/N$. In other situations a more delicate study of the posterior distribution involving more than the variance may be necessary.

REFERENCES

HALDANE, J. B. S. (1945). *Biometrika*, **33**, 222–5.
JEFFREYS, H. (1948). *Theory of Probability*, 2nd ed. Oxford: Clarendon Press.
LINDLEY, D. V. (1956). *Ann. Math. Statist.* **27**, 986–1005.
LINDLEY, D. V. (1957). *J. R. Statist. Soc.* B (to appear).
MAULDON, J. G. (1955). *J. R. Statist. Soc.* B, **17**, 79–85.
STEIN, C. (1955). *Ann. Math. Statist.* **26**, 157.

A STATISTICAL PARADOX

By D. V. LINDLEY

Statistical Laboratory, University of Cambridge

An example is produced to show that, if H is a simple hypothesis and x the result of an experiment, the following two phenomena can occur simultaneously:

(i) a significance test for H reveals that x is significant at, say, the 5 % level;

(ii) the posterior probability of H, given x, is, for quite small prior probabilities of H, as high as 95 %.

Clearly the common-sense interpretations of (i) and (ii) are in direct conflict. The phenomenon is fairly general with significance tests and casts doubts on the meaning of a significance level in some circumstances.

We begin by giving the mathematical derivation of the example and later comment on it and the assumptions involved. Let $(x_1, x_2, \ldots, x_n)$ be a random sample from a normal distribution of mean θ and known variance σ^2. Let the prior probability that $\theta = \theta_0$, the value on the null hypothesis, be c. Suppose that the remainder of the prior probability is distributed uniformly over some interval I containing θ_0. We shall deal with situations where $\bar{x}$, the arithmetic mean of the observations, and a minimal sufficient statistic, is well within the interval I. The posterior probability that $\theta = \theta_0$, in the light of the sample, can be evaluated; it is

$$\bar{c} = c \exp[-n(\bar{x}-\theta_0)^2/(2\sigma^2)]/K, \tag{1}$$

where

$$K = c \exp[-n(\bar{x}-\theta_0)^2/(2\sigma^2)] + (1-c)\int_I \exp[-n(\bar{x}-\theta)^2/(2\sigma^2)]\,d\theta,$$

by Bayes's theorem. In virtue of the assumption about $\bar{x}$ and I the integral can be evaluated as $\sigma\sqrt{(2\pi/n)}$.

Now suppose that the value of $\bar{x}$ is such that, on performing the usual significance test for the mean θ_0 of a normal distribution with known variance, the result is significant at the α percentage point. That is, $\bar{x} = \theta_0 + \lambda_\alpha \sigma/\sqrt{n}$, where λ_α is a number dependent on α only and can be found from tables of the normal distribution function. Inserting this value for $\bar{x}$ in (1) we have the following value for the posterior probability that $\theta = \theta_0$

$$\bar{c} = c\, e^{-\frac{1}{2}\lambda_\alpha^2} / \{c\, e^{-\frac{1}{2}\lambda_\alpha^2} + (1-c)\,\sigma\sqrt{(2\pi/n)}\}. \tag{2}$$

(Note that $\bar{x}-\theta_0$ tends to zero as n increases so that $\bar{x}$ will lie well within the interval I for sufficiently large n.) From (2) we see that as $n \to \infty$, $\bar{c} \to 1$. It follows that whatever the value of c, a value n can be found, dependent on c and α such that

(i) $\bar{x}$ is significantly different from θ_0 at the α % level;

(ii) the posterior probability that $\theta = \theta_0$ is $(100-\alpha)$ %.

This is the paradox. The usual interpretation of the first result is that there is good reason to believe $\theta \neq \theta_0$; and of the second, that there is good reason to believe $\theta = \theta_0$. The two interpretations are in direct conflict, and the conflict may apparently be made even stronger by remarking that the $(100-\alpha)$ % confidence and fiducial intervals for θ just exclude $\theta = \theta_0$. With $\alpha = 5$ we are 95 % confident that $\theta \neq \theta_0$, but have 95 % belief that $\theta = \theta_0$.

In commenting on this analysis, let us first consider the assumptions involved. Many significance tests involve situations in which the test criterion is asymptotically normally

distributed with known variance, as is $\bar{x}$ in the example, and therefore the sample considered is in no way unusual. The only assumption that will be questioned is the assignment of a prior distribution of any type, and, in particular, of the form chosen. A paradox will only have been generated if we can show there exist situations where (*a*) a prior distribution of this form is reasonable, and (*b*) a significance test of the 'tail-area' type is commonly used. Let us first consider the assignment of any prior probability. The argument for the use of prior probabilities has been put forward very cogently by Jeffreys (1948). His arguments have, to my mind, been reinforced by those of Ramsey (1931) and, more especially, Savage (1954). Savage's main contribution is as follows: he lays down certain axioms that a man should follow if he is to act in a 'rational' way, and defines a rational man to be a man who acts according to these axioms. The latter are quite mild in their form and would surely be agreed to by most statisticians. Savage then shows that a rational man must act as if he had a prior probability distribution and (if relevant) a utility function. It does not follow from this that any statistical inference need make overt mention of a prior distribution, but it does follow that no inference procedure should grossly contradict the existence of a prior distribution. (A mild contradiction may be allowable in the interests of simplicity.) Another way of looking at this result is to say that a probability distribution is a satisfactory measure of one's convictions about several hypotheses. For example, if to-day we say that our prior belief in one hypothesis is $\frac{1}{2}$ it will mean the same as saying to-morrow that our prior belief in a different hypothesis is $\frac{1}{2}$; just as a yard of material to-day measures the same as a yard of material to-morrow. If we are to use a significance level in a similar way, as Fisher (1956, p. 43) has suggested we can, and most statisticians do, we must establish a similar comparison property. 5% to-day must mean the same as 5% to-morrow. Our example, we claim, shows that it need not.

So much for the general question of introducing a prior distribution. We now consider the particular form used in deriving the paradox. We first note that the phenomenon would persist with almost any prior probability distribution that had a concentration on the null value and no concentrations elsewhere. For example, if there is an amount c at $\theta = \theta_0$ and the rest is distributed throughout I according to a density $p(\theta)$, where $\int_I p(\theta)\,d\theta = 1-c$, then if $p(\theta)$ is bounded it is easy to show, for example, by a steepest descent argument applied to the integral corresponding to that in (1), that $\bar{c}$ still tends to 1. It is sufficient that $p(\theta)$ does not tend to infinity too rapidly as θ tends to θ_0. It is, however, essential that the concentration on the null value exists, and it is this that has to be considered. Again Jeffreys (1948) has discussed the point. Briefly, one argument is that the singling out of the hypothesis $\theta = \theta_0$ to be tested is itself evidence that the value θ_0 is in some way special and is likely therefore to be true. We should like to give two examples where this seems unquestionably correct. The first is in genetics where θ is the linkage parameter between two genetic factors. If there is no linkage $\theta = \theta_0 = \frac{1}{4}$, and we are concerned with developing a test to determine if there is any evidence for linkage. Now in this situation there is a considerable amount of prior knowledge. For it is known that there is linkage if, and only if, the two genes lie on the same chromosome. Consequently if there are n chromosomes of approximately equal length, and if it seems reasonable to suppose that the gene is equally likely to be anywhere along the chromosomes' lengths, then it seems reasonable to suppose a prior probability of the order of $(n-1)/n$ that the value of θ is $\frac{1}{4}$. The particular numerical value of the prior probability is not so important here (though we note it is rather large)

as is the fact that $\theta = \frac{1}{4}$ is in a singular position and will arise for most positions of the genes. A second example arises in the telepathy experiments carried out by Soal & Bateman (1954), where, if no telepathic powers are present, the experiment has a success ratio of $\theta = \frac{1}{5}$, otherwise $\theta \neq \frac{1}{5}$. A significance test for telepathy therefore should assign to $\theta_0 = \frac{1}{5}$ a concentration of probability equal to one's prior belief that the subject has not got telepathic powers. This example is perhaps not as convincing as the genetical one because of the prejudices that exist in connexion with extra-sensory perception. My point in both these examples is that the value θ_0 is fundamentally different from any value of $\theta \neq \theta_0$, however near to θ_0 it might be. Unquestionably there exist situations (perhaps they are the more common) in which this is not so; where we are interested in testing the approximate validity of the null hypothesis, such as that the treatment has no (or very little) effect. This point has been discussed by Hodges & Lehmann (1954).

We now consider the paradox in these situations where the prior probability exists (by Savage's argument) and has a concentration on the null value. We first note that the expression of it in terms of fiducial or confidence limits used above is unjustified. The limits purport to be statements made about the value of θ in the light of the experimental result when initially nothing is known about or independent of knowledge of θ. The type of prior distribution used here (suggested by the practical circumstances of the problem) certainly does not correspond to ignorance about θ. Thus we should not be surprised at the disagreement. The paradox merely serves as a warning that the confidence or fiducial type of statement should only be used in those circumstances where one is truly ignorant about the parameter. We have argued that this is not so in the telepathy or genetical examples.

The conflict between statements of a significance level and statements based on Bayes's theorem remains. Now in our example we have taken situations in which the significance level is fixed because, as explained above, we wish to see whether its interpretation as a measure of lack of conviction about the null hypothesis does mean the same in different circumstances. The Bayesian probability is all right, by the arguments above; and since we now see that it varies strikingly with n for fixed significance level, in an extreme case producing a result in direct conflict with the significance level, the degree of conviction is not even approximately the same in two situations with equal significance levels. 5% in to-day's small sample does not mean the same as 5% in to-morrow's large one.

An alternative interpretation of the paradox was suggested to me by Prof. Barnard. The posterior probability $\bar{c}$, given by (2), may be written

$$\bar{c} = cf_n/\{cf_n + (1-c)\},$$

where

$$f_n = \sqrt{\left(\frac{n}{2\pi\sigma^2}\right)} e^{-\frac{1}{2}\lambda_z^2},$$

the likelihood of θ_0 on the evidence of the sample. Clearly $f_n \to \infty$ as $n \to \infty$, λ_z fixed. Hence for fixed significance level the likelihood of the null hypothesis increases indefinitely with the sample size. This appears to me to demonstrate, without reference to prior probabilities, the unsoundness of the suggestion that significance tests depend on the disjunction: either a rare chance has occurred or the null hypothesis is false (Fisher, 1956, p. 39). For the chance considered in a significance test is the chance of the observed event and other more extreme ones. The chance of the observed event is measured by the likelihood function. These two chances behave quite differently. In fact, the paradox arises because the significance level

argument is based on the area under a curve and the Bayesian argument is based on the ordinate of the curve. However, the above interpretation through the likelihood involves no mention of alternative hypotheses which seem basic to any approach to the problem.

The other approach to significance testing, due to Neyman & Pearson, does envisage the use of alternative hypotheses and hence appears to give a reason for using the tail area because this region is the best one in which to reject the null hypothesis at a specified level of significance. Therefore the occurrence of an observation in the region is an unusual event on the null hypothesis and less unusual on some alternative hypotheses. But the theory does not justify the practice of keeping the significance level fixed, nor does it take account of the fact that when the observation has been made we know, not that the point has fallen in the region of significance, but that it has fallen exactly on the edge, and the likelihoods under the null and alternative hypotheses seem the relevant quantities to compare.

The paradox is not, in essentials, new, although few statisticians are aware of it. The difference between the two approaches has been noted before by Jeffreys (see, in particular, 1948, Appendix), who is the originator of significance tests based on Bayes's theorem and a concentration of prior probability on the null value. But Jeffreys is concerned to emphasize the similarity between his tests and those due to Fisher and the discrepancies are not emphasized. The same phenomenon was noticed by Lindley (1953) in decision theory studies, and some computations by Prof. Pearson in the discussion to that paper emphasized how the significance level would have to change with the sample size, if the losses and prior probabilities were kept fixed. (The discussion based only on the latter quantities is mathematically equivalent to one in decision theory language with zero-one losses.) The present note considers the situation where the significance level is fixed and the variation in posterior probability is evaluated, rather than the other way round.

The concept of a significance level has been used very successfully in practical problems of inference. One might now ask how this has come about. The answer has already been given by Jeffreys in the appendix already cited. Essentially it is because $\bar{c}$, as given by (2), tends to unity very slowly and, for moderate values of n, $\bar{c}$ may be less than c at a prescribed significance level and the two concepts be in reasonable agreement. Let

$$A = c\,e^{-\frac{1}{2}\lambda_\alpha^2}/(1-c)\sqrt{(2\pi)},$$

then

$$\bar{c} = A/(A+\sigma/\sqrt{n}), \tag{3}$$

and $\bar{c} \to 0$ as $\sigma/\sqrt{n} \to \infty$. Hence in a small experiment, significance at 5 % may give very strong reasons to doubt the null hypothesis. A numerical example is informative. Suppose we take $c = \frac{1}{2}$ and use a two-sided test at 5 % significance so that $\lambda_\alpha = 1{\cdot}96$; then $A = 0{\cdot}0584$ and the table gives the value of $\bar{c}$ for different values of $t = n/\sigma^2$. If $\sigma = 1$, $t = n$, and we see that for small samples ($n \leqslant 10$) the probability of θ_0 has decreased appreciably from its initial value of $\frac{1}{2}$, giving cause to doubt the validity of the null hypothesis. For medium samples ($10 < n < 100$) the probability has only decreased a little, so that although we are not as confident as we were initially about the null hypothesis, our doubts are not great. By the time n has reached a value about 300 $\bar{c}$ is equal to c; the experiment, despite its 5 % significance, has not altered our belief in the null hypothesis at all. To reach the strong contrast put forward in the paradox it would be necessary to take n about 10,000. Of course if σ is smaller then smaller samples will suffice. For example, if we apply these numerical values to the Soal & Bateman problem (i.e. use the normal approximation to the binomial) we have $\sigma^2 = \frac{1}{5}.\frac{4}{5} = 0{\cdot}16$, and a sample of size about 48 has $\bar{c}$ equal to the

original value of $\frac{1}{2}$. An experiment of this type with a run of forty-eight trials which is significant at 5% would not alter our views on telepathy if initially we had an open mind on the problem. The normal approximation is not adequate for samples as low as 10, but it is clear that only such small ones would increase our prior belief at all noticeably. An experiment of 1600 trials would raise our belief that telepathy did not exist to 95%; quite a moderate size in comparison with the 37,100 trials carried out with Mrs Stewart. The reader may be interested to know that with $c = \frac{1}{2}$ the posterior probability of the null hypothesis $\theta = \frac{1}{5}$ in the light of the experiments with Mrs Stewart (9410 successes) is of the order of 10^{-140}. The evidence for Mrs Stewart's telepathic powers is rather strong.

t	$\bar{c}$	t	$\bar{c}$
1	0·055	600	0·589
2	0·076	800	0·623
3	0·092	1,000	0·649
4	0·105	2,000	0·723
5	0·116	4,000	0·787
10	0·156	6,000	0·819
20	0·207	8,000	0·839
40	0·270	10,000	0·854
60	0·312	20,000	0·892
80	0·343	40,000	0·921
100	0·369	60,000	0·935
200	0·453	80,000	0·943
300	0·503	100,000	0·949
400	0·539	∞	1·000

An apparent advantage of the significance level statement is that it does provide some sort of assessment of the truth of the null hypothesis using only the evidence provided by the experiment. It is, in effect, a convenient (though possibly misleading) summary of what the experimental result has to say about the null hypothesis. A similar assessment is available in a Bayesian analysis through the likelihood function. In the situation considered here the function is proportional to

$$\sqrt{\left(\frac{n}{2\pi}\right)} \exp\{-\tfrac{1}{2}n(\bar{x}-\theta)^2\},$$

regarded as a function of θ. This, unlike the single number expressing the significance level, is a function and is therefore more difficult to understand. A reduction to a numerical value is possible provided the assessment of prior probabilities conditional on $\theta \neq \theta_0$ is made. For example, if θ is uniform in the interval I in these circumstances, then

$$\exp[-n(\bar{x}-\theta_0)^2/2\sigma^2] \Big/ \int_I \exp[-n(\bar{x}-\theta)^2/2\sigma^2]\,d\theta = \sqrt{\left(\frac{n}{2\pi}\right)} \exp\left\{-\frac{n(\bar{x}-\theta_0)^2}{2\sigma^2}\right\}$$

is the quantity by which the prior odds, $c/(1-c)$, in favour of θ_0 must be multiplied in order to obtain the posterior odds, $\bar{c}/(1-\bar{c})$. This single value, or its logarithm, might be an acceptable substitute for the significance level. It is numerically equal to Jeffreys's K, since he supposes $c = \frac{1}{2}$.

The paradox serves to explain one puzzling feature of tests based on Bayes's theorem. Suppose the experimenter has continued sampling randomly until he has reached a result which is, using a fixed-sample size significance test, significant at some prescribed significance level α. That is, he has taken a sample $(x_1, x_2, \ldots, x_n)$ such that $\bar{x} = \theta_0 + \lambda_\alpha \sigma/\sqrt{n}$.

It is easy to show, by the law of the iterated logarithm, that this will happen with probability one whatever the value of θ. Then the experimenter has, of course, cheated if he quotes his result as being significant at $\alpha\%$, though, if the distribution theory were known, a valid significance test could be made. But it would not be that appropriate to a sample of fixed size n. On the other hand, it is easy to see that the likelihood of the observations $(x_1, x_2, \ldots, x_n)$ does not depend on the particular sequential stopping rule used and is, therefore, equal to the likelihood the experimenter would have obtained if the same sample had been reached by taking a sample of fixed size n. It follows that any significance test based on Bayes's theorem does not depend on the sequential stopping rule used, at least amongst a wide class of such rules. In the extreme case the experimenter can go on sampling until he has reached the significance level α, and yet the fact that he did so is irrelevant to a Bayesian. In telepathy this is known as 'optional stopping': stopping when the results look striking; striking, that is, on a significance level criterion. The explanation is now clear. If $\theta \neq \theta_0$ the optional stopper will reach his desired point for small n and $\bar{c} < c$. On the other hand, if $\theta = \theta_0$ the value of n will be larger and $\bar{c} > c$. (These are average results, of course, naturally sometimes mistakes will be made.) The value of $\bar{c}$ is just what one would expect in the two cases and we see that the Bayesian will not on the average be in error in ignoring the stopping rule. It should now be possible to give a reliable assessment of those results in telepathy which have had objections raised against them on the grounds of optional stopping.

I am much indebted to Profs. Pearson and Barnard for helpful comments on the first draft of this paper.

REFERENCES

FISHER, R. A. (1956). *Statistical Methods and Scientific Inference.* Edinburgh: Oliver and Boyd.
HODGES, J. L. & LEHMANN, E. L. (1954). *J. R. Statist. Soc.* B, **16**, 261–8.
JEFFREYS, H. (1948). *Theory of Probability*, 2nd ed. Oxford: Clarendon Press.
LINDLEY, D. V. (1953). *J. R. Statist. Soc.* B, **15**, 30–76.
RAMSEY, F. P. (1931). *The Foundations of Mathematics.* London: Routledge and Kegan Paul.
SAVAGE, L. J. (1954). *The Foundations of Statistics.* New York: Wiley.
SOAL, S. G. & BATEMAN, F. (1954). *Modern Experiments in Telepathy.* London: Faber and Faber.

STOCHASTIC CROSS-INFECTION BETWEEN TWO OTHERWISE ISOLATED GROUPS

By H. W. HASKEY

*South-east Essex Technical College**

1. Bailey (1950) and Haskey (1954) have considered, from a stochastic standpoint, the spread of a mild infection among a single group of people into which one infectious individual was introduced, the group being isolated. The following is a generalization: There are now two groups of people, one with n_1 individuals and the other with n_2. At time $t=0$ we have h infectives introduced into the first group. At time t we suppose there are, respectively, r_1 and r_2 uninfected susceptibles remaining. Let the chance of one of the susceptibles in the first group being infected in time dt by an infective in the same group be $ar_1(n_1+h-r_1)\,dt$; and by an infective in the second group be $b'r_1(n_2-r_2)\,dt$. The total chance of a fresh infection is then the sum of these quantities. Similarly, for the susceptibles in the second group, let the chances of infection from infectives of the same or of the other group be $br_2(n_2-r_2)\,dt$ and $a'r_2(n_1+h-r_1)\,dt$. Writing $p(r_1, r_2)$ for the probability of r_1, r_2 susceptibles at time t and considering the possible changes,

$$(r_1+1, r_2)\to(r_1, r_2), \quad (r_1, r_2+1)\to(r_1, r_2), \quad (r_1, r_2)\to(r_1, r_2)$$

in the ensuing time dt, we have, events in the groups being regarded as independent,

$$\begin{aligned}\frac{dp(r_1, r_2)}{dt} = &-p(r_1, r_2)\,[ar_1(n_1+h-r_1)+br_2(n_2-r_2)+a'r_2(n_1+h-r_1)+b'r_1(n_2-r_2)]\\ &+p(r_1, r_2+1)\,[b(r_2+1)\,(n_2-r_2-1)+a'(r_2+1)\,(n_1+h-r_1)]\\ &+p(r_1+1, r_2)\,[a(r_1+1)\,(n_1+h-r_1-1)+b'(r_1+1)\,(n_2-r_2)]. \qquad (1)\end{aligned}$$

If the Laplace transform of $p(r_1, r_2)$ with respect to time is $q_\lambda(r_1, r_2)$, or $q(r_1, r_2)$ for brevity, we have

$$\begin{aligned}q(r_1, r_2)\,[\lambda+f(r_1, r_2)] = &(r_2+1)\,[b(n_2-r_2-1)+a'(n_1+h-r_1)]\,q(r_1, r_2+1)\\ &+(r_1+1)\,[a(n_1+h-r_1-1)+b'(n_2-r_2)]\,q(r_1+1, r_2), \qquad (2)\end{aligned}$$

where

$$f(r_1, r_2) = (ar_1+a'r_2)\,(n_1+h-r_1)+(br_2+b'r_1)\,(n_2-r_2). \qquad (3)$$

This holds if $r_2<n_2$, $r_1<n_1$. Likewise

$$q(r_1, n_2)\,[\lambda+f(r_1, n_2)] = a(r_1+1)\,(n_1+h-r_1-1)\,q(r_1+1, n_2) \quad (r_1<n_1), \qquad (4)$$

$$q(n_1, r_2)\,[\lambda+f(n_1, r_2)] = (r_2+1)\,[b(n_2-r_2-1)+a'h]\,q(n_1, r_2+1) \quad (r_2<n_2), \qquad (5)$$

$$q(n_1, n_2)\,[\lambda+f(n_1, n_2)] = 1. \qquad (6)$$

2. To find the mean number uninfected in the group originally uninfected, the quantities $m(r_1, r_2) = r_2q(r_1, r_2)$ are required. From (2) it follows that

$$m(r_1, r_2)\,[\lambda+f(r_1, r_2)] = j(r_1, r_2+1)\,m(r_1, r_2+1)+k(r_1+1, r_2)\,m(r_1+1, r_2), \qquad (7)$$

with similar equations corresponding to (4), (5) and (6), where

$$\left.\begin{aligned}j(r_1, r_2) &= (r_2-1)\,[b(n_2-r_2)+a'(n_1+h-r_1)],\\ k(r_1, r_2) &= r_1[a(n_1+h-r_1)+b'(n_2-r_2)].\end{aligned}\right\} \qquad (8)$$

* Now at U.K. Atomic Energy Authority (Industrial Group H.Q.)

To see the nature of the solution we take $n_1=2$, $n_2=3$ and work backwards from $m(2,3)$ to $m(0,1)$, giving

$$\left.\begin{aligned} &m(2,3) = 3/[f(2,3)+\lambda], \quad m(1,3) = 3k(2,3)/[f(2,3)+\lambda][f(1,3)+\lambda],\\ &m(0,3) = 3k(2,3)\,k(1,3)/[f(2,3)+\lambda][f(1,3)+\lambda][f(0,3)+\lambda],\\ &m(2,2) = 3j(2,3)/[f(2,3)+\lambda][f(2,2)+\lambda],\\ &m(1,2) = 3k(2,3)j(1,3)/[f(2,3)+\lambda][f(1,3)+\lambda][f(1,2)+\lambda]\\ &\qquad\qquad + 3j(2,3)\,k(2,2)/[f(2,3)+\lambda][f(2,2)+\lambda][f(1,2)+\lambda] \quad \text{etc.} \end{aligned}\right\} \tag{9}$$

Any $m(r_1, r_2)$ can be written from the table below:

r_1 \ r_2	3	2	1
2			
1			
0			

by selecting all 'approved routes' from the cell $(2,3)$ to the cell (r_1, r_2). An 'approved route' starts from $(2,3)$, runs along rows to the right and down columns. Corresponding to each approved route to (r_1, r_2) there is a term of $m(r_1, r_2)\,[f(2,3)+\lambda]/3$ consisting of a product of fractions of the types

$$j(s_1, s_2)/[f(s_1, s_2-1)+\lambda] = \theta(s_1, s_2), \quad \text{say},$$

for motion to the right and

$$k(s_1, s_2)/[f(s_1-1, s_2)+\lambda] = \phi(s_1, s_2), \quad \text{say},$$

for motion down a column, from the cell (s_1, s_2).

There are as many terms to be added to give $m(r_1, r_2)$ as there are different approved routes from the cell $(2,3)$ to the cell (r_1, r_2). The double sum $\sum_{r_1}\sum_{r_2} m(r_1, r_2)$, required to find the mean number of uninfected in the group originally uninfected, is the sum of contributions from all approved routes to every cell. These contributions can be arranged in sets, each set being associated with one approved route to the cell $(0,1)$. Hence

$$\sum_{r_1}\sum_{r_2} m(r_1, r_2) = 3\psi(2,3)/[f(2,3)+\lambda], \tag{10}$$

where $\psi(j,k)$ is defined inductively by

$$\psi(j,k) = 1+\theta(j,k)\,\psi(j,k-1)+\phi(j,k)\,\psi(j-1,k) \tag{11}$$

and $\psi(0,1) = \psi(1,0) = 1$. Equation (10) can be expanded and the double sum consists of products like (9).

When $n_1 = 2$, $n_2 = 3$, $h = 3$ and $a = 0{\cdot}8$, $b = 0{\cdot}9$, $a' = 0{\cdot}2$, $b' = 0{\cdot}2$ and the value of $f(r_1, r_2)$ at each cell is inserted within it, we have

r_1 \ r_2	3	2	1
2	6·6	8·2	8·0
1	5·6	6·8	6·2
0	3·0	3·8	2·8

Each term of the double sum has then to be resolved into partial fractions, becoming $\sum_{r_1}\sum_{r_2} a_{r_1, r_2}/[f(r_1, r_2)+\lambda]$, and the different terms collected; for example, $a_{23}/(6{\cdot}6+\lambda)$, $a_{22}/(8{\cdot}2+\lambda)$, and the coefficients a_{r_1, r_2} computed. Corresponding to these we get inverse Laplace transforms $a_{23}e^{-6{\cdot}6t}$, etc., and their sum gives the mean susceptibles M at time t. Then $-dM/dt$ is the number of 'cases' at time t.

The above is a relatively simple case—though the computations are lengthy—since there are no two values of $f(r_1, r_2)$ alike. When there are repeated values, terms involving $(k+\lambda)^{-1}$, $(k+\lambda)^{-2}$, $(k+\lambda)^{-3}$, etc., appear in the double sum and the calculation of their coefficients is involved. For the numerical case cited above we obtain

	Term in mean number infected	Term of 'cases'
$f(2, 3)$	$-\ 9{\cdot}8791e^{-6{\cdot}6t}$	$-\ 65{\cdot}202e^{-6{\cdot}6t}$
$f(1, 3)$	$-\ 4{\cdot}9934e^{-5{\cdot}6t}$	$-\ 27{\cdot}963e^{-5{\cdot}6t}$
$f(0, 3)$	$-\ 99{\cdot}6923e^{-3{\cdot}0t}$	$-\ 299{\cdot}077e^{-3{\cdot}0t}$
$f(2, 2)$	$0{\cdot}6787e^{-3{\cdot}2t}$	$5{\cdot}565e^{-3{\cdot}2t}$
$f(1, 2)$	$3{\cdot}5457e^{-2{\cdot}8t}$	$24{\cdot}111e^{-2{\cdot}8t}$
$f(0, 2)$	$14{\cdot}3491e^{-2{\cdot}8t}$	$54{\cdot}526e^{-2{\cdot}8t}$
$f(2, 1)$	$-\ 0{\cdot}8242e^{-3{\cdot}0t}$	$-\ 6{\cdot}593e^{-3{\cdot}0t}$
$f(1, 1)$	$10{\cdot}8647e^{-2{\cdot}2t}$	$67{\cdot}361e^{-2{\cdot}2t}$
$f(0, 1)$	$88{\cdot}9412e^{-2{\cdot}8t}$	$249{\cdot}035e^{-2{\cdot}8t}$

The sum of the terms in the first column at $t = 0$ is 3, as it should be, and the sum of the terms in the second column for various values of t gives cases as follows:

t	0	0·1	0·3	0·5	0·6	1·0
Cases	1·76	2·21	2·67	2·54	2·37	1·44

Thus we have a case curve which rises steadily to a maximum and then decays to zero as time elapses.

3. I have been unable to obtain a general formula for $\Sigma\Sigma m(r_1, r_2)$ in the general case, since this expression involves sums of terms like those in a hypergeometric series except that the factors in these terms are quadratic in r_1, r_2 and in general without factors.

While the mathematical model of infection used here is likely to be approximately true in general, there are good epidemiological grounds for thinking that it applies only to small

groups for which n_1, n_2 do not exceed five, and that instead of two cross-infecting groups there are many. Even for two groups of four the full expression for $\Sigma\Sigma m(r_1, r_2)$ in the contracted notation of equation (10) is a yard long. Leaving aside the question of three cross-infecting groups—which would involve a three-dimensional group of cells instead of the two-dimensional group in §2—there is a special case for which a general formula can be obtained for the mean susceptibles in either group for any values of n_1, n_2. This arises when a, b, a', b' are equal. We imagine constant fractions of the infectives and the uninfected susceptibles in each group travelling to the other for a given infinitesimal interval. They then return to their own group for another different given infinitesimal interval and follow this by another visit to the other group and so on. This gives (1) but depends on the doubtful procedure of splitting infinitesimal time intervals, but if it is granted, then the special case corresponds to the state of affairs in which only the susceptibles, and all of them, travel to and fro, spending half their time in each group, or alternatively when only the infectives, and all of them, travel, spending equal times in each group. Choosing the unit of time to make $a = 1$, we consider this special case.

For simplicity taking $h = 1$, and $n_2 = n_1 + 1 = n$, $f(r_1, r_2)$ factorizes into $(r_1 + r_2)(2n - r_1 - r_2)$ and the table for $f(r_1, r_2)$ is:

r_1 \ r_2	n	$n-1$		2	1
$n-1$	$(2n-1).1$	$(2n-2).2$		$(n+1).(n-1)$	$n.n$
$n-2$	$(2n-2).2$	$(2n-3).3$		$n.n$	$(n-1).(n+1)$
0	$n.n$	$(n-1).(n+1)$		$2.(2n-2)$	$1.(2n-1)$

This shows that there are many repeated entries but that any approved route from the beginning of the epidemic, $(n-1, n)$, to its end, $(0, 1)$, passes through each of the different possible values of $f(r_1, r_2)$ twice and twice only.

We now write $x(2n-x) = \Omega(x) = \mathbf{x}$, say, with the understanding that whenever several clarendon symbols are connected by signs the complete set of them forms the argument for Ω. Thus $\mathbf{x} - \mathbf{y} = \Omega(x-y)$, not $\Omega(x) - \Omega(y)$. It then follows that

$$f(r_1, r_2) = \mathbf{r}_1 + \mathbf{r}_2,$$

$$(\mathbf{2n - s - m - 1} + \lambda)\,\theta(n-m, n-s) = (n-s-1)(s+m),$$

$$(\mathbf{2n - s - m - 1} + \lambda)\,\phi(n-m, n-s) = (n-m)(s+m),$$

and
$$\Sigma\Sigma m(r_1, r_2) = \frac{n}{\mathbf{2n-1}+\lambda}\left[1 + \frac{(n-1)}{\mathbf{2n-2}+\lambda}\{\psi(n-1, n-1) + \psi(n-2, n)\}\right],$$

and by repeated use of

$$\psi(r_1, r_2) = 1 + \theta(r_1, r_2)\,\psi(r_1, r_2 - 1) + \phi(r_1, r_2)\,\psi(r_1 - 1, r_2),$$

we find that

$$\Sigma\Sigma m(r_1, r_2) = \frac{n}{\mathbf{2n-1}+\lambda}\left\{1+\frac{\{2n-2\}.1}{\mathbf{2n-2}+\lambda}\right.$$
$$\left.\times\left[1+\frac{\{2n-3\}.2}{\mathbf{2n-3}+\lambda}\left(1+\ldots+\frac{2.\{2n-3\}}{\mathbf{2}+\lambda}\left[1+\frac{1.\{2n-2\}}{\mathbf{1}+\lambda}\right]\ldots\right)\right]\right\}. \quad (12)$$

Since $\mathbf{x} = \mathbf{2n}-\mathbf{x}$, all such terms as $\mathbf{4}+\lambda$ occur twice in denominators, except $\mathbf{n}+\lambda$, which occurs once only. The coefficients of $(\mathbf{r}+\lambda)^{-1}$ and $(\mathbf{r}+\lambda)^{-2}$ in the expansion of (12) in partial fractions must be found, and we can, without loss of generality, consider all values of $r<n$ and then the isolated value $r=n$.

All the terms containing $(\mathbf{r}+\lambda)^{-2}$ and not merely $(\mathbf{r}+\lambda)^{-1}$ are included in T_r, where $T_r = A_r B_r$ and

$$\left.\begin{aligned} A_r &= \frac{n(2n-2)!\,(2n-r-1)!}{(r-1)!\,(\mathbf{1}+\lambda)\,(\mathbf{2}+\lambda)\ldots(\mathbf{r-1}+\lambda)\,[(\mathbf{r}+\lambda)\,(\mathbf{r+1}+\lambda)\ldots(\mathbf{n-1}+\lambda)]^2\,(\mathbf{n}+\lambda)}, \\ B_r &= \left\{1+\frac{(r-1)\,(2n-r)}{\mathbf{r-1}+\lambda}\left[1+\frac{(r-2)\,(2n-r+1)}{\mathbf{r-2}+\lambda}\left(1+\ldots+\frac{2.(2n-3)}{\mathbf{2}+\lambda}\left[1+\frac{1.(2n-2)}{\mathbf{1}+\lambda}\right]\ldots\right)\right]\right\}, \end{aligned}\right\} \quad (13)$$

and in partial fractions T_r is

$$c_{r1}/(\mathbf{1}+\lambda)+\ldots+c_{rr}/(\mathbf{r}+\lambda)+\ldots+c_{rn}/(\mathbf{n}+\lambda)+d_{r1}/(\mathbf{1}+\lambda)^2+\ldots+d_{r,n-1}/(\mathbf{n-1}+\lambda)^2, \quad (14)$$

where
$$d_{rr} = T_r(\mathbf{r}+\lambda)^2|_{\lambda=-r}, \quad c_{rr} = \frac{\partial}{\partial\lambda}\{T_r(\mathbf{r}+\lambda)^2\}|_{\lambda=-r}, \quad \text{etc.} \quad (15)$$

Noting that $n^{(r)} = n(n-1)\ldots(n-r+1)$, the series B_r becomes

$$1+\frac{(-)^1(r-1)^{(1)}(r-2n)^{(1)}}{\mathbf{r-1}+\lambda}+\frac{(-)^2(r-1)^{(2)}(r-2n)^{(2)}}{(\mathbf{r-1}+\lambda)\,(\mathbf{r-2}+\lambda)}+\ldots \text{ to } r \text{ terms.} \quad (16)$$

It is found that

$$(\lambda+\mathbf{r})^2 A_r|_{\lambda=-r} = (-)^{r-1}\,2n(n-r)\,(2n-2)!/[(r-1)!]^2\,(2n-2r-1)! \quad [r=1,2,\ldots,(n-1)], \quad (17)$$

and
$$B_r|_{\lambda=-r} = 1+\frac{(-)^1(r-1)^{(1)}(r-2n)^{(1)}}{1!\,(2r-2n-1)^{(1)}}+\frac{(-)^2(r-1)^{(2)}(r-2n)^{(2)}}{2!\,(2r-2n-1)^{(2)}}+\ldots \text{ to } r \text{ terms}$$
$$= F(1-r, 2n-r; 2n-2r+1; 1)$$
$$= (-)^{r-1}(2n-2r)!\,(r-1)!/(2n-r-1)!. \quad (18)$$

Hence
$$d_{rr} = 4n(n-r)^2\,(2n-2)!/(r-1)!\,(2n-r-1)!, \quad (19)$$

further
$$\left.\begin{aligned} c_{rr} &= \left[B_r\frac{\partial}{\partial\lambda}\{A_r(\lambda+\mathbf{r})^2\}+A_r(\lambda+\mathbf{r})^2\frac{\partial B_r}{\partial\lambda}\right]\Big|_{\lambda=-r}, \\ \frac{\partial}{\partial\lambda}[A_r(\lambda+\mathbf{r})^2]\,(r-1)!\,(\mathbf{1}+\lambda)\,(\mathbf{2}+\lambda)\ldots(\mathbf{r-1}+\lambda) & \\ \times[(\mathbf{r}+\mathbf{1}+\lambda)\,(\mathbf{r}+\mathbf{2}+\lambda)\ldots(\mathbf{n-1}+\lambda)]^2\,(\mathbf{n}+\lambda) & \\ = -n(2n-2)!\,(2n-r-1)!\left\{\sum_{k=1}^{r-1}1/(\lambda+\mathbf{k})+1/(\lambda+\mathbf{n})+2\sum_{k=r+1}^{n-1}1/(\lambda+\mathbf{k})\right\}. & \end{aligned}\right\} \quad (20)$$

Setting $\lambda = -\mathbf{r}$ it is found that

$$[(r-1)!]^2\,(2n-2r-1)!\left[\frac{\partial}{\partial\lambda}A_r(\lambda+\mathbf{r})^2\right]_{\lambda=-r}$$
$$= (-)^r\,n(2n-2)!\left\{-\sum_{z=1}^{r-1}z^{-1}+\sum_{2n-2r+1}^{2n-r-1}z^{-1}+2\sum_{z=1}^{2n-2r-1}z^{-1}\right\}. \quad (21)$$

If $r=1$, the first two sums of reciprocals in (21) are omitted.

Since the evaluation of $\partial B_r/\partial\lambda$ by differentiation of each term of B_r above leads to an awkward expression, the general term of B_r, viz.

$$(-)^x (r-1)^{(x)} (r-2n)^{(x)}/(\mathbf{r-1}+\lambda)(\mathbf{r-2}+\lambda)\dots(\mathbf{r-s+1}+\lambda)(\mathbf{r-x}+\lambda) \quad (r-1\geqslant x\geqslant s), \tag{22}$$

may be put into partial fractions as follows

$$\sum_{s=1}^{x} \frac{2(-)^x (r-1)^{(x)} (r-2n)^{(x)} (r-n-s)}{(-)^{s-1}(s-1)!\,(x-s)!\,(2r-1-2n-s)^{(x)}} \frac{1}{(\mathbf{r-s}+x)}. \tag{23}$$

On differentiating this with respect to λ and putting $\lambda=-\mathbf{r}$, there results

$$\sum_{s=1}^{x} \frac{2(-)^{x-s}(r-1)^{(x)}(r-2n)^{(x)}(r-n-s)}{ss!\,(x-s)!\,(2r-2n-s+1)^{(x)}(2r-2n-s)^2} = \sum_{s=1}^{x} g(x,s), \quad \text{say,} \tag{24}$$

so that

$$\left.\frac{\partial B_r}{\partial\lambda}\right|_{\lambda=-\mathbf{r}} = \sum_{x=1}^{r-1}\sum_{s=1}^{x} g(x,s). \tag{25}$$

This double sum may be expressed as $\sum_{s=1}^{r-1}\sum_{x=s}^{r-1} g(x,s)$, or as

$$\sum_{s=1}^{r-1} \frac{(-)^s 2(r-n-s)}{ss!\,(2r-2n-s)^2}\left\{\frac{(-)^s(r-1)^{(s)}(r-2n)^{(s)}}{0!\,(2r-2n-s-1)^{(s)}} + \frac{(-)^{s+1}(r-1)^{(s+1)}(r-2n)^{(s+1)}}{1!\,(2r-2n-s-1)^{(s+1)}}\right.$$
$$\left. + \dots + \frac{(-)^{r-1}(r-1)^{(r-1)}(r-2n)^{(r-1)}}{(r-1-s)!\,(2r-2n-s-1)^{(r-1)}}\right)$$
$$= \sum_{s=1}^{r-1} \frac{2(r-n-s)(r-1)^{(s)}(r-2n)^{(s)}}{ss!\,(2r-2n-s)^2(2r-2n-s-1)^{(s)}}\left(\sum_{y=0}^{y=r-1-s}\frac{(-)^y(r-2n-s)^{(y)}(r-1-s)^{(y)}}{y!\,(2r-2n-1-2s)^{(y)}}\right)$$
$$= \sum_{s=1}^{r-1} \frac{2(r-n-s)(r-1)^{(s)}(r-2n)^{(s)}F(s+1-r, 2n-r+s;\ 2n+2s-2r+1;\ 1)}{ss!\,(2r-2n-s)^2(2r-2n-s-1)^{(s)}}$$
$$= \sum_{s=1}^{r-1} 2(r-n-s)(r-1)!\,(r-2n)^{(s)}/(2r-2n-2s-1)^{(r-1-s)}(2r-2n-s-1)^{(s)}(2r-2n-s)^2 ss!.$$

Since $$(2r-2n-s-1)^{(r-1)} = (2r-2n-s-1)^{(s)}(2r-2n-2s-1)^{(r-1-s)},$$

the above becomes

$$2(r-1)!\sum_{s=1}^{r-1}(r-n-s)(r-2n)^{(s)}/(2r-2n-s-1)^{(r-1)}(2r-2n-s)^2 ss!.$$

Further, as

$$(2r-2n-1)^{(r-1)}(r-2n)^{(s)} = (2r-2n-1)^{(s+r-1)} = (2r-2n-1)^{(s)}(2r-2n-s-1)^{(r-1)},$$

we have

$$(2r-2n-1)^{(r-1)}\left.\frac{\partial B_r}{\partial\lambda}\right|_{\lambda=-\mathbf{r}} = 2(r-1)!\sum_{s=1}^{r-1}(2r-2n-1)^{(s-1)}(r-n-s)/s(2r-2n-s)\,s! \quad (r>2). \tag{26}$$

If $r=2$, $$\left.\frac{\partial B_2}{\partial\lambda}\right|_{\lambda=-\mathbf{2}} = -2(n-1)/(2n-3)^2.$$

Expressing $2(r-n-s)/s(2r-2n-s)$ as $1/s+1/(p+s)$,

where $p=2n-2r$, we have

$$\left.\frac{\partial B_r}{\partial\lambda}\right|_{\lambda=-\mathbf{r}} = \sum_{s=1}^{r-1}\{1/s+1/(p+s)\}(-)^{r-s}(r-1)(r-2)\dots(s+1)/(p+s)(p+s+1)\dots(p+r-1).$$

If this is written out as two series, putting the terms in the order for which $s = r-1, r-2, \ldots$, there results

$$\left.\frac{\partial B_r}{\partial\lambda}\right|_{\lambda=-r} = -\frac{1}{(p+r-1)}\left[\frac{1}{r-1} - \frac{(r-1)}{(r-2)(p+r-2)} + \frac{(r-1)(r-2)}{(r-3)(p+r-2)(p+r-3)} \quad \text{to } (r-1) \text{ terms}\right.$$

$$\left. + \frac{1}{p+r-1} - \frac{(r-1)}{(p+r-2)^2} + \frac{(r-1)(r-2)}{(p+r-3)^2(p+r-2)} \quad \text{to } (r-1) \text{ terms}\right]$$

$$= -(p+r-1)^{-1}\left[\int_0^1 [(x^{p+r-2}+x^{r-2})\{F_l(1-r, 1; -p-r+2; -x^{-1})-1\}]\,dx + (r+1)^{-1} + (p+r-1)^{-1}\right], \tag{27}$$

where the suffix l in the hypergeometric indicates that its last term is to be omitted. If $r=2$ the integral is to be omitted and if $r=1$, $\partial B_r/\partial\lambda = 0$.

All the terms in $\Sigma\Sigma m(r_1, r_2)$ containing an unrepeated factor $(\mathbf{r}+\lambda)^{-1}$ are included in u_r, where

$$= \frac{n(2n-2)!\,(r-1)!}{(\mathbf{2n-1}+\lambda)(\mathbf{2n-2}+\lambda)\ldots(\mathbf{2n-r}+\lambda)(2n-r-1)!}$$

$$\times\left\{1+\frac{(2n-r-1)r}{(\mathbf{2n-r-1}+\lambda)}\left[1+\frac{(2n-r-2)(r+1)}{(\mathbf{2n-r-2}+\lambda)}\left(1+\ldots\frac{(r+2)(2n-r-3)}{\mathbf{r+2}+\lambda}\left[1+\frac{(r+1)(2n-r-2)}{\mathbf{r+1}+\lambda}\right]\ldots\right)\right]\right\} \quad (r \leqslant n), \tag{28}$$

and the coefficient of $(\mathbf{r}+\lambda)^{-1}$ in this is $u_r(\lambda+\mathbf{r})\,|_{\lambda=-\mathbf{r}}$, which simplifies to

$$(-)^{r-1}n(2n-2)!\,(2n-2r)!\,F(r, r+1-2n;\ 2r+1-2n;\ 1)/[(2n-r-1)!]^2 \quad (r = 1, 2, \ldots, n). \tag{29}$$

Gathering the terms and taking the inverse Laplace transforms, the mean number of susceptibles in the group originally uninfected at time t is

$$\sum_{r=1}^{n}\left[(-)^{r-1}n(2n-2)!\,(2n-2r)!\,F(r, r+1-2n;\ 2r+1-2n;\ 1)/[(2n-r-1)!]^2\right.$$

$$-2n(n-r)(2n-2)!\left\{-\sum_{z=1}^{r-1}z^{-1} + \sum_{z=2n-2r+1}^{2n-r-1}z^{-1} + 2\sum_{1}^{2n-2r+1}z^{-1}\right\}\Big/(2n-r-1)!\,(r-1)!$$

$$+\frac{(-)^r 2n(n-r)(2n-2)!}{[(r-1)!]^2(2n-2r-1)!\,(2n-r-1)}$$

$$\times\left\{\int_0^1[(x^{2n-r-2}+x^{r-2})\{F_l(1-r, 1; 2-p-r;\ -x^{-1})-1\}]\,dx + (r-1)^{-1} + (2n-r-1)^{-1}\right\}$$

$$\left.+4n(n-r)^2(2n-2)!\,t/(r-1)!\,(2n-r-1)!\right]\exp[-r(2n-r)t], \tag{30}$$

from which the mean number of cases can be obtained. In computations it is better to work with (12) and use (30) as a check.

4. The method of finding the mean number of susceptibles in the group, numbering n_1, is similar to that in § 2, for equation (1) holds and new quantities $n(r_1, r_2) = r_1 q(r_1, r_2)$

$$j'(r_1, r_2) = r_2\{b(n_2-r_2)+a'(n_1+h-r_1)\},$$
$$k'(r_1, r_2) = (r_1-1)\{a(n_1+h-r_1)+b'(n_2-r_2)\},$$

are defined, the primes indicating that they relate to the place of origin of the epidemic. The procedure is as before, but now r_2 runs from n_2 to 0 and r_1 from n_1 to 1.

If $n_2=3$, $n_1=2$, $h=1$, $a=\frac{3}{4}$, $b=\frac{2}{3}$, $a'=\frac{1}{6}$, $b'=\frac{1}{2}$, we have

Table of $f(r_1, r_2)$

$r_2\rightarrow$	3	2	1	0
r_1	2	4	$7\frac{1}{3}$	8	6
$\downarrow$	1	5	$7\frac{1}{2}$	$7\frac{1}{3}$	$4\frac{1}{2}$

Table of $j'(r_1, r_2)$

$r_2\rightarrow$	3	2	1	0
r_1	2	1	$3\frac{1}{3}$	3	0
$\downarrow$	1	2	4	$3\frac{1}{3}$	0

Table of $k'(r_1, r_2)$

$r_2\rightarrow$	3	2	1	0
r_1	2	$1\frac{1}{2}$	2	$2\frac{1}{2}$	3
$\downarrow$	1	0	0	0	0

Hence

$$\sum_{r_1}\sum_{r_2} n(r_1, r_2) = 2\psi(2,3)/[f(2,3)+\lambda] = 2[1+\theta(2,3)\,\psi(2,2)+\phi(2,3)\,\psi(1,3)]/[f(2,3)+\lambda] = \text{etc.},$$

giving

$$\Sigma\Sigma n(r_1, r_2) = \frac{2}{4+\lambda}\left\{1+\frac{1}{7\frac{1}{3}+\lambda}\left[1+\frac{3\frac{1}{3}}{8+\lambda}\left(1+\frac{3}{6+\lambda}\overline{1+\frac{3}{4\frac{1}{2}+\lambda}}\right.\right.\right.$$

$$\left.\left.\left.+\frac{2\frac{1}{2}}{7\frac{1}{3}+\lambda}\overline{1+\frac{3\frac{1}{3}}{4\frac{1}{2}+\lambda}}\right)\right]+\frac{3}{5+\lambda}+\frac{1}{7\frac{1}{3}+\lambda}\left(\frac{2}{7\frac{1}{3}+\lambda}+\frac{3}{5+\lambda}\right)\left(1+\frac{4}{7\frac{1}{3}+\lambda}\overline{1+\frac{3\frac{1}{3}}{4\frac{1}{2}+\lambda}}\right)\right\},$$

The coefficients of $1/(7\frac{1}{3}+\lambda)$, $1/(7\frac{1}{3}+\lambda)^2$ in the above are those of $e^{-7\frac{1}{3}t}$ and $te^{-7\frac{1}{3}t}$ in the expression M for the mean number of susceptibles, which is

$$M = 35{\cdot}2079e^{-4t} - 54{\cdot}0979e^{-4\frac{1}{2}t} + 17{\cdot}9143e^{-5t} + 3{\cdot}75e^{-6t}$$
$$+ 27{\cdot}6571e^{-7\frac{1}{2}t} + 1{\cdot}5179e^{-8t} + (6{\cdot}4059t - 29{\cdot}9493)\,e^{-7\frac{1}{3}t}, \qquad (31)$$

yielding the following table:

Cases $= -dM/dt$	3	2·84	2·52	2·08	1·46	0·71	0·04
t	0	0·18	0·30	0·42	0·60	0·90	1·80

(32)

The cases show a steady decline with elapse of time. For the group to which the epidemic originally spreads when $n_2=3$, $n_1=2$, $h=1$, $a=\frac{3}{4}$, $b=\frac{2}{3}$, $a'=\frac{1}{6}$, $b'=\frac{1}{2}$, the cases are given by a formula similar to (31) yielding the result:

Cases	1	2·299	2·523	2·476	1·916	0·469	0·02
t	0	0·3	0·54	0·6	0·9	1·2	3·0

(33)

It is seen that the cases in this group rise slowly to a maximum at about time $t=0{\cdot}55$ and then decline slowly to zero.

When the cases in both groups are totalled for given times the case-time curve is similar to that of a self-infecting community, rising to a maximum and declining rapidly thereafter. The maximum total cases occur at about time $=0{\cdot}24$, at which time the cases for a single self-infecting group of the same total size is a maximum, but this self-infecting group has a higher maximum than the total cases in the two cross-infecting groups. The same results have been found for different values of a, b, a', b', except when $a=b=a'=b'$, when the total cases in two cross-infecting groups is the same as for a single self-infecting group of the same total size.

5. The individual probabilities $p(r_1, r_2)$ can be calculated by the foregoing method. This is illustrated for the special case, taking $n_2 = 3 = n_1+1$ and considering the group originally

uninfected. $\Sigma q(r_1, r_2)$ now includes the quantity $\phi(1,0)$, since $\sum_r p(r_1, 0)$ is found from $q(2,0)+q(1,0)+q(0,0)$. It is found that

$$\sum_{r_1} p(r_1, 0) = 1+(29-48t)\,e^{-5t}+(7-84t)\,e^{-8t}-37e^{-9t}, \tag{a}$$

$$\Sigma p(r_1, 1) = (-37+48t)\,e^{-5t}+(120t-20)\,e^{-8t}+57e^{-9t}, \tag{b}$$

$$\Sigma p(r_1, 2) = 6e^{-5t}+(15-36t)\,e^{-8t}-21e^{-9t}, \tag{c}$$

$$\Sigma p(r_1, 3) = 2e^{-5t}-2e^{-8t}+e^{-9t}. \tag{d}$$

If these probabilities are plotted against time on the same axes it is seen that (a) rises steadily from zero to unity while (d) decreases steadily from unity to zero. (b) and (c) rise from zero to maxima of about 0·3, 0·32 respectively, attained at times 0·5, 0·24 respectively, and then decrease to zero.

At any time t the graph which has the greatest ordinate shows the most probable number of susceptibles, ζ, and it is found that

Time	0–0·32	0·32–0·37	0·37–0·45	0·45 onwards
ζ	3	2	1	0
Approx. prob. at transition times	0·3	0·278	0·295	

Observational data would be obtained in this form.

6. A convenient expression for the total mean number of susceptibles in the two groups can also be obtained when $f(r_1, r_2)$ factorizes. By way of illustration we now write $\Lambda(x) = x(n_1+n_2+h-x)$ as $\mathbf{x}$, then

$$f(r_1, r_2) = (r_1+r_2)\,(n_1+n_2+h-r_1-r_2) = \mathbf{r}_1+\mathbf{r}_2.$$

The Laplace transform of the total mean susceptibles in the groups is

$$\Sigma\Sigma l(r_1, r_2) = \Sigma\Sigma(r_1+r_2)\,q(r_1, r_2),$$

and writing $N = n_1+n_2+h$,

$$\begin{aligned}(r_1+r_2)\,j''(r_1, r_2) &= r_2(r_1+r_2-1)\,(N-r_1-r_2),\\ (r_1+r_2)\,k''(r_1, r_2) &= r_1(r_1+r_2-1)\,(N-r_1-r_2),\end{aligned} \tag{34}$$

there results

$$[f(r_1, r_2)+\lambda]\,l(r_1, r_2) = l(r_1, r_2+1)\,j''(r_1, r_2+1)+l(r_1+1, r_2)\,k''(r_1+1, r_2). \tag{35}$$

The table for $f(r_1, r_2)$ is

r_1 ↓ \ r_2 →	n_2	n_2-1	n_2-2	...
n_1	$h(n_1+n_2)$	$(h+1)\,(n_1+n_2-1)$	$(h+2)\,(n_1+n_2-2)$	...
n_1-1	$(h+1)\,(n_1+n_2-1)$	$(h+2)\,(n_1+n_2-2)$	...	...
n_1-2	$(h+2)\,(n_1+n_2-2)$	...	...	...
...	...	...	...	...

in which all entries in diagonals running from top right to bottom left are the same. The solution of the difference equation (35) gives, by the usual method,

$$\begin{aligned}[\lambda+f(n_1, n_2)]\,\Sigma\Sigma l(r_1, r_2)/(n_1+n_2) &= \psi(n_1, n_2)\\ &= 1+\theta(n_1, n_2)\,\psi(n_1, n_2-1)+\phi(n_1, n_2)\,\psi(n_1-1, n_2),\end{aligned}$$

where $\theta(n_1, n_2) = j''(n_1, n_2)/[f(n_1, n_2-1)+\lambda]$, $\phi(n_1, n_2) = k''(n_1, n_2)/[f(n_1-1, n_2)+\lambda]$.

Inserting the values of $\psi(n_1, n_2-1)$, $\psi(n_1-1, n_2)$ and of their successors there results

$$\Sigma\Sigma l(r_1, r_2) = \frac{n_1+n_2}{\mathbf{N-h}+\lambda}\left[1+\frac{1}{\mathbf{N-h-1}+\lambda}\left\{(N-h-1)\,h+\frac{1}{\mathbf{N-h-2}+\lambda}\right.\right.$$
$$\left.\left.\times\left\{(N-h-1)\,(N-h-2)\,h(h+1)+\frac{1}{\mathbf{N-h-3}+\lambda}\cdots\right.\right.\right.$$

continued until the term $\mathbf{1}+\lambda$ appears in the denominator. If $n_1=n_2=n$ and $h=1$, remembering that now $\mathbf{x} = (\mathbf{2n+1-x})$, we find that

$$\Sigma\Sigma l(r_1, r_2) = 2n\left[\frac{1}{\lambda+\mathbf{1}}+\frac{(2n-1)!\,1!}{(2n-2)!\,(\lambda+\mathbf{1})(\lambda+\mathbf{2})}+\ldots+\frac{(2n-1)!\,(p-1)!}{(2n-p)!\,(\lambda+\mathbf{1})(\lambda+\mathbf{2})\ldots(\lambda+\mathbf{p})}\right.$$
$$\left.+\ldots+\frac{\{(2n-1)!\}^2}{[(\lambda+\mathbf{1})(\lambda+\mathbf{2})\ldots(\lambda+\mathbf{n})]^2}\right]. \quad (36)$$

If the differential-difference equations for a single self-infecting community of $2n$ persons and one infectious person are written down (Bailey's equation (5) with n replaced by $2n$) and $P(r)$ replaces $rp(r)$, we have

$$\frac{dP(r)}{dt} = r(2n-r)\,P(r+1)-r(2n+1-r)\,P(r),$$
$$\frac{dP(2n)}{dt} = -2nP(2n).$$

Taking the Laplace transforms and noting

$$\mathbf{r} = r(2n+1-r) = \mathbf{2n+1-r},$$
$$\left.\begin{aligned}(\mathbf{r}+\lambda)\,Q(r) &= r(2n-r)\,Q(r+1)\\ (2n+\lambda)\,Q(2n) &= 2n\end{aligned}\right\} \quad (r=0, 1, \ldots, (2n-1)), \quad (37)$$

whence

$$\sum_r Q(r) = 2n\left[\frac{1}{\lambda+\mathbf{1}}+\frac{(2n-1)!}{(2n-2)!\,(\lambda+\mathbf{1})(\lambda+\mathbf{2})}+\frac{(2n-1)!\,2!}{(2n-3)!\,(\lambda+\mathbf{1})(\lambda+\mathbf{2})(\lambda+\mathbf{3})}+\ldots\right]. \quad (38)$$

From (36) and (38) it is seen that $\Sigma l(r_1, r_2)$ is identical with $\sum_r Q(r)$. Hence the solution, obtained by the present writer for $\Sigma Q(r)$ in a self-infecting group, can be used for this case of cross-infection. Hence the total mean susceptibles in the two cross-infecting groups is

$$\sum_{r=1}^{n}\frac{2n!\,(2n+1-2r)}{(r-1)!\,(2n-r)!}\left\{2/(2n+1-2r)-\sum_{z=r}^{2n-r}z^{-1}+(2n+1-2r)\,t\right\}\exp\left[-r(2n+1-r)\,t\right]. \quad (39)$$

7. If n in (39) is large, computations to find the number of cases at any time are lengthy, and even if n is small the general formula for cases, resembling (31), is not informative of the nature of the graphs M, $-dM/dt$ against t. In this section it is proposed to develop an approximation to the total case curve for the two cross-infecting groups, considered in § 6. From (36)

$$Q(r) = 2n!\,(2n-r)!\,[(1+\mathbf{r}/\lambda)\,(1+\{\mathbf{r}+\mathbf{1}\}/\lambda)\ldots(1+\mathbf{2n}/\lambda)]^{-1}/(r-1)!\,\lambda^{2n+1-r}, \quad (40)$$

and the Laplace transform of the total mean number of susceptibles is $\sum_0^{2n} Q(r)$ or

$$2n[1+\mathbf{2n}/\lambda]^{-1}/\lambda+2n(2n-1)\,[(1+\mathbf{2n}/\lambda)\,(1+\{\mathbf{2n-1}\}/\lambda)]^{-1}/\lambda^2$$
$$+2n(2n-1)\,(2n-2)\,2!\,[(1+\mathbf{2n}/\lambda)\,(1+\{\mathbf{2n-1}\}/\lambda)\,(1+\{\mathbf{2n-2}\}/\lambda)]^{-1}/\lambda^3+\ldots. \quad (41)$$

Since λ is greater than any of the $\mathbf{r}$, the binomials may be expanded and the coefficients of powers of λ^{-1} in them are the sums of homogeneous products of appropriate dimensions of one or more of the quantities $\mathbf{2n}$, $\mathbf{2n-1}$, $\mathbf{2n-2}$, etc. For the first few powers of λ^{-1} we have

$$\sum_0^{2n} Q(r) = 2n[\lambda^{-1} - \lambda^{-2} - 2(n-1)\lambda^{-3} - 4(n^2-4n+2)\lambda^{-4} + 8(-n^3+11n^2-19n+7)\lambda^{-5} - 16(n^4-26n^3+107n^2-123n+38)+\ldots]. \quad (42)$$

As λ^{-n} has an inverse Laplace transform with respect to time of $t^{n-1}/(n-1)!$, the mean susceptibles at time t is

$$2n[1-t-(n-1)t^2-2(n^2-4n+2)t^3/3+t^4(-n^3+11n^2-19n+7)/3 - 2t^5(n^4-26n^3+107n^2-123n+38)/15+\ldots], \quad (43)$$

and mean cases are

$$2n[1+2(n-1)t+2(n^2-4n+2)t^2-4t^3(-n^3+11n^2-19n+7)/3 + 2t^4(n^4-26n^3+107n^2-123n+38)/3\ldots]. \quad (44)$$

This is quicker than the method given by Bailey (equations (15)–(23)), but only slowly convergent, and a practical alternative to the expansion of the exponentials, $e^{-\mathbf{rt}}$, with their formidable coefficients.

It has been found that the case curve cannot be approximated to satisfactorily by one or even two curves of the type $(a+bt)e^{-kt}$ of which it is composed. Since it appears to be nearly a truncated normal curve, the first three terms of (44) were supposed to commence the expansion of $k\exp[-a(t-b)^2]$ giving $a = 4n-2$, $b = (n-1)/(4n-2)$ which can then be written

$$2n\exp\{(n-1)^2/(4n-2)\}\exp\{-(4n-2)[t-(n-1)/(4n-2)]^2\} \times\{1-4t^3(n-3)(2n-1)/3-2t^4(2n-1)(n^2-11n+16)/3+\ldots\}, \quad (45)$$

where the series, producing skewness, must be continued to many terms.

For a given numerical value of n a reasonably good approximation to the case curve can be obtained by the method of Karl Pearson. When $n = 2, 3$ the mean cases are from (39),

$$(-23+36t)e^{-4t}+(24+18t)e^{-6t}$$

and

$$(1140+720t)e^{-12t}+(2700t-645)e^{-10t}+(900t-489)e^{-6t},$$

respectively.

The single negative values of t for which these curves cut the t-axis are $-0{\cdot}3918$ and $-0{\cdot}2500$ respectively. The areas under the curves and above the t-axis from these points to infinity are easily found, the cases at time t are divided by the respective areas and finally the origin is shifted to the points of intersection of the curve with the t-axis.

For $n=2$, $T = t+0{\cdot}3918$ and the equation of the adjusted case curve is

$$\gamma(T) = (-146{\cdot}91+142{\cdot}59T)e^{-4T}+(146{\cdot}91+156{\cdot}08T)e^{-6T},$$

which is like a probability density function in that $\int_0^\infty \gamma(T)\,dT = 1$.

The first four moments of $\gamma(T)$ about $T=0$ and the mean, $T=0{\cdot}7984$, are

$$\mu_2' = 0{\cdot}8370, \quad \mu_3' = 1{\cdot}0679, \quad \mu_4' = 1{\cdot}5880$$

and

$$\mu_2 = 0{\cdot}1996, \quad \mu_3 = 0{\cdot}0809, \quad \mu_4 = 0{\cdot}1597.$$

These yield $\beta_1 = \mu_3^2/\mu_2^3 = 0{\cdot}823, \quad \beta_2 = \mu_4/\mu_2^2 = 4{\cdot}01,$

which, by Biometrika Table 43, indicates the Pearson Type I curve as the best fit, although Type III might fit better for higher values of n. For $n=2$ the cases at time t are

$$\text{const.} \times (1+t/0{\cdot}3918)^{1{\cdot}64}\,(1-t/b_2)^{17}, \quad b_2 \doteq 5{\cdot}88.$$

This gives the mode at time 0·16, rather earlier than the correct value 0·21, obtained by solving $\gamma'(t) = 0$.

Using T in the sense above and using a Type III curve it is found that

$$\gamma(T) = e^{-4T}\,T^{2{\cdot}193} \quad \text{for } n=2, \quad \gamma(T) = e^{-5{\cdot}96T}\,T^{2{\cdot}555} \quad \text{for } n=3.$$

It is surmised in general that $\gamma(T) \doteq e^{-2nT}\,T^{2+p}$, where p is positive. An improved fit could be obtained by calculation of exact case values, using (39), for selected times and applying maximum likelihood to calculate the constants.

REFERENCES

BAILEY, N. T. J. (1950). *Biometrika*, **37**, 193.
HASKEY, H. W. (1954). *Biometrika*, **41**, 272.

SOME STATISTICS ASSOCIATED WITH THE RANDOM DISORIENTATION OF CUBES

By J. K. MACKENZIE and M. J. THOMSON

Division of Tribophysics, Commonwealth Scientific Industrial Research Organization, University of Melbourne, Australia

ABSTRACT. A Monte Carlo method is used to estimate the frequency functions of various angles in a random aggregate of cubic crystals. Estimates are made of the frequency function for the angle of disorientation, i.e. the least angle of rotation required to rotate a crystal into the same orientation as a neighbouring crystal, and for the angles Min ⟨100⟩, Min ⟨110⟩, Min ⟨112⟩, Min ⟨123⟩ and Min [⟨110⟩, ⟨112⟩, ⟨123⟩], where Min ⟨100⟩ is defined as the least of the nine acute angles between ⟨100⟩ directions in neighbouring crystals and similar definitions apply for the other angles.

1. Introduction

This paper is concerned with the estimation, by means of random sampling, of some probability distributions which arise from a class of problems in three-dimensional geometrical probability. The results obtained are of interest to metallurgists in particular and perhaps to crystallographers in general. They are presented here in the hope that some statistician may be sufficiently interested to try to obtain the exact distributions in a more or less explicit form.

Before stating the specific problems it is desirable in the interests of the general reader to describe the standard crystallographic notation for directions and planes. A particular direction can be specified by the components u, v, w of a vector (in this direction) relative to an orthonormal basis and the symbol $[uvw]$, in square brackets, is used to denote this direction. Thus, [100] is the direction of the x-axis. A cube with its centre at the origin and edges parallel to the base vectors (axes) is invariant under the 48 symmetry operations of the cubic group consisting of 24 proper rotations and 24 improper rotations which are proper rotations together with an inversion or reflexion. Starting with a given direction $[uvw]$, 47 other equivalent directions (24 lines in all) can be derived by the use of the symmetry operations and are called variants of $[uvw]$. These are simply derived from the given direction by permuting the indices u, v, w in sign and in order in all possible ways. This set of 48 equivalent directions is denoted by $\langle uvw \rangle$, in carets, and not all the 48 directions need be distinct, e.g. the set ⟨100⟩ consists of the set of 6 directions [100], $[\bar{1}00]$, [010], $[0\bar{1}0]$, [001], $[00\bar{1}]$, the bar over an index being used instead of a minus sign. Similarly, if h, k, l are the components of a normal to a particular plane, this plane is denoted by the symbol (hkl), in brackets. While the set of all planes equivalent to the plane (hkl) is derived in the same way and denoted by $\{hkl\}$, in braces. These symbols for planes are not used in the present paper.

The simplest of all the problems under consideration can be stated as follows. Given a single fixed reference line (defined by one of two opposite directions) and another single line defined by a random direction, uniformly distributed on a sphere, what is the probability distribution of the least angle between these two lines or of the least angle of rotation required to make the random line coincide with the reference line? It is known that the cosine of both these angles is uniformly distributed in the range (0, 1), but what is the answer if instead of two single lines there are two congruent (i.e. superposable) sets of lines,

the lines of each set being fixed relative to one another? The present paper gives practical answers to some problems of this latter type when the lines of each set are invariant under the rotations of the cubic group.

2. Statement of problems

Consider two cubes, A and B, and imagine A to be a fixed reference cube and B to be initially coincident with A but free to rotate in any manner about the common centre of A and B. If B is rotated through an arbitrary angle about some arbitrary axis there are 24 definite rotations which will restore B into coincidence with A. These rotations are just the reverse of the original rotation taken together with the 24 proper symmetry operations associated with a cube having indistinguishable faces. Further, each of these 24 rotations can be represented as a single rotation about some definite axis and through some definite angle. Then, of the 24 angles of rotation so defined, there is one (or more) which is least in magnitude and this least angle may be taken as a measure of the disorientation of the two cubes and will be called the angle of disorientation.

In 1949, F. C. Frank proposed over morning coffee the problem of determining the greatest possible angle of disorientation of two cubes. The answer to the analogous problem for squares in two dimensions is of course 45°, but in three dimensions the answer is not at all obvious. However, by a tedious consideration of all possibilities it can be shown that the maximum value of the angle of disorientation is $2 \arccos \frac{1}{4}(2+\sqrt{2}) = 62{\cdot}80°$, and that the rotation which achieves this maximum disorientation is most simply described as a rotation of 90° about any of the axes $\langle 110 \rangle$, i.e. axes parallel to a face diagonal. A more difficult problem is to determine the probability distribution of the angle of disorientation when the cube B takes all orientations with equal probability. An estimate of this probability distribution is made in the present paper.

Another problem which has arisen in the course of experimental work (Ogilvie, 1952) can be described in simplified form as follows. In an aggregate of cubic crystals a particular event may occur only when one of the directions, $\langle 100 \rangle$ say, in one crystal is within, say, 5° of one of the $\langle 100 \rangle$ directions in a neighbouring crystal. Then it may be asked what proportion of pairs of crystals in a random aggregate would comply with this last requirement. The problem can be formulated as follows. Imagine a set of three fixed wires passing through the centre and parallel to the directions $\langle 100 \rangle$ of cube A and a similar set of three wires parallel to the directions $\langle 100 \rangle$ of cube B. Then, if B is given a random rotation there are nine definite acute angles between pairs of wires (taking one from each set) and the least of these angles will be called Min $\langle 100 \rangle$. The probability distributions of Min $\langle 100 \rangle$, and the analogous Min $\langle 110 \rangle$, Min $\langle 112 \rangle$, Min $\langle 123 \rangle$ and Min [$\langle 110 \rangle$, $\langle 112 \rangle$, $\langle 123 \rangle$] are also estimated in the present paper.

The method of calculation is described in § 3 and the results are given in § 4. The method of constructing 150 random orthogonal matrices and their testing for randomness is set out in § 5.

3. Method of calculation

The calculations were performed by the method of random sampling. Since any rotation can be represented by a 3×3 orthogonal matrix (Jeffreys & Jeffreys, 1946, p. 114), 150 random orthogonal 3×3 matrices were constructed as described in § 4. Then with each matrix the following calculations were made.

If a given matrix represents a rotation through an angle θ about some axis, then the trace of the matrix (i.e. the sum of the elements in the leading diagonal) is equal to $1+2\cos\theta$. Thus, to determine the angle of disorientation it is necessary, in principle, to calculate the trace of each of the 24 matrices found by combining the given matrix with the 24 symmetry operations of a cube and to choose the greatest of these 24 values. However, this greatest trace can be found easily from the following rule. Taking into account only the magnitude of each element, add to the largest element the greater of the two diagonal sums of the four elements not lying in the same row or column as the largest element. This rule can be derived by straightforward but tedious consideration of all the possibilities.

Since the elements of the given matrix are just the cosines of the 9 angles between the new and the old $\langle 100\rangle$ directions, Min $\langle 100\rangle$ is determined by the element of largest magnitude in the matrix.

The remaining calculations are all similar and only the determination of Min $\langle 123\rangle$ will be described. First the matrix is used to calculate the new directions corresponding to each of the 24 (proper) variants of [123] (the remaining 24 are obtained by changing all signs). If a particular member of the set $\langle 123\rangle$ is transformed into $[u_1u_2u_3]$ then the cosine of the angle between this direction and the nearest direction of the set $\langle 123\rangle$ is

$$\tfrac{1}{14}(|u_s|+2|u_m|+3|u_l|),$$

where u_s, u_m and u_l are the numerically smallest, middle and largest of u_1, u_2 and u_3. The greatest of the 24 cosines so calculated determines Min $\langle 123\rangle$. A table showing the bounds for $|u_s|$, $|u_m|$ and $|u_l|$ consistent with a series of different angular deviations from [123] was used to reject most of the 24 possibilities by visual inspection; the cosines were only computed accurately for the few remaining cases.

Finally, when the results of the calculations had been accumulated for the 150 matrices, the number of cases in which the various angles lay in suitably chosen ranges were counted and the numbers so obtained used as estimates of the corresponding frequencies.

4. RESULTS AND DISCUSSION

The results are presented in the form of a series of histograms in Fig. 1. The ordinates have been normalized to represent probability densities when the unit of measurement along the abscissa is 1°, and the figures along the top of each histogram are the actual number of cases counted in the indicated range. If p is the estimated probability of an angle lying in a particular range, then an estimate of the standard error of p based on a sample of 150 is $[p(1-p)/150]^{\frac{1}{2}}$, and horizontal dotted lines have been drawn one standard error above and below the top of each rectangle of the histograms. The mean $\bar{x}$ and the standard deviation s of the estimated distribution are given on each histogram and are also indicated by means of the arrow and range at the bottom of each histogram.

The dotted curves superposed on each histogram give an indication of the form of each frequency function. These have been adjusted so that the area under each curve is unity. Except for Min $\langle 110\rangle$ each frequency function appears to have a single maximum, and even in the case of Min $\langle 110\rangle$ the existence of a double hump is by no means certain. However, there are, for finer subdivisions of the ranges, indications of more than one hump in some of the other frequency functions and the nature of the problem suggests that the true frequency functions may consist of a number of continuous arcs which join at sharp corners.

As was mentioned in the introduction, a tedious argument shows that the greatest possible value for the angle of disorientation is 62·80°. A similar argument shows that the greatest possible value of Min ⟨100⟩ is $\arccos\frac{2}{3} = 48{\cdot}19°$ and that the rotation which achieves the corresponding disorientation is most simply described as a rotation of 60° about an axis ⟨111⟩. Although no further results of this type are known to the authors it is easily

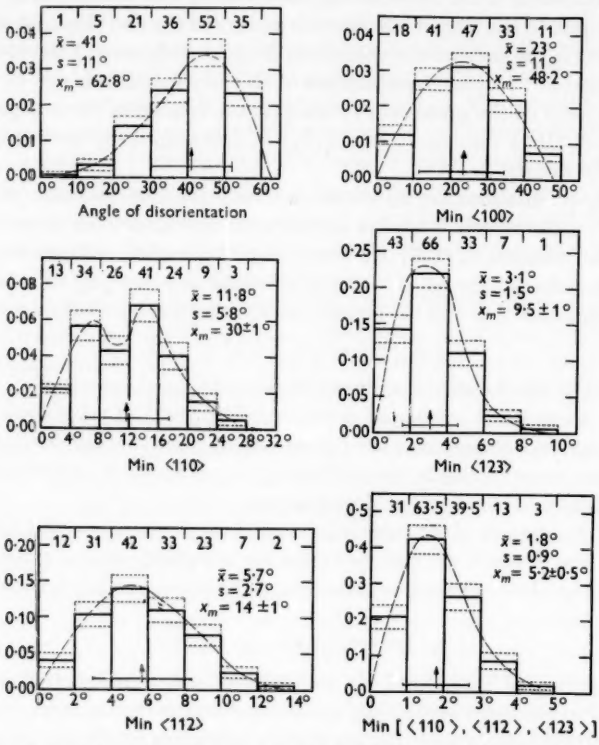


Fig. 1. Histograms derived from a random sample of 150. The ordinate is probability density when the angles are measured in degrees. The figures at the top are the number of cases counted in each range while the horizontal dotted lines indicate limits corresponding to one standard error from the estimated total probability for each range. The mean $\bar{x}$ and the standard deviation s of the estimated distribution are indicated by the arrow and range at the bottom. An estimate of the maximum value x_m of each variable is also given. The dotted curves indicate the general shape of each frequency function.

shown that a rotation of 45° about an axis ⟨100⟩ gives Min ⟨110⟩ = $\arccos\frac{1}{4}(2+\sqrt{2}) = 31{\cdot}40°$, and the results given in Fig. 1 suggest that this is probably its greatest possible value. The values, x_m, of the greatest possible values of each variable are given in Fig. 1 and where the value is associated with limits of error it has been estimated as follows.

A rotation is uniquely defined by three suitable independent variables (e.g. the Eulerian angles) so that the variation of Min ⟨110⟩, say, can be represented by means of a four-dimensional hypersurface. Now experience obtained in the calculation of the maximum values of the angle of disorientation and Min ⟨100⟩ suggests that in all cases the appropriate

hypersurface can be approximated in the neighbourhood of $x = x_m$ by a number of hyperplanes all intersecting at a common point. Thus, when x is sufficiently near x_m the probability density would be expected to be proportional to $(x_m - x)^2$. Except for the angle of disorientation and Min $\langle 100 \rangle$ this expected behaviour can be roughly verified from Fig. 1 and x_m has been estimated by plotting $p^{\frac{1}{2}}$ against the mean value of x for the range concerned. The limits of error stated in Fig. 1 are no more than reasonable guesses and the maximum $30 \pm 1°$ given for Min $\langle 110 \rangle$ is to be compared with the known result that the maximum is probably 31·40°.

A simple argument accounts for the initial linear rise of the frequency functions for all the Min $\langle uvw \rangle$ in Fig. 1. If the set $\langle uvw \rangle$ has $2n$ members defining n distinct lines Min $\langle uvw \rangle$ is the least of the n^2 angles θ_i between pairs of lines and $\cos\theta_i$ is uniformly distributed on the range (0, 1). Now the method of inclusion and exclusion shows that $\Pr\{\text{Min}\langle uvw\rangle < \theta\}$ lies between the limits* $\sum_i \Pr\{\theta_i < \theta\}$ and $\sum_i \Pr\{\theta_i < \theta\} - \sum_{i<j} \Pr\{\theta_i < \theta, \theta_j < \theta\}$. Thus, for θ sufficiently small

$$\Pr\{\text{Min}\langle uvw\rangle < \theta\} \simeq n^2 \Pr\{\theta_i < \theta\} = n^2(1 - \cos\theta).$$

The corresponding density function is $n^2 \sin\theta$, and the estimated frequency functions in Fig. 1 are in substantial agreement with this prediction up to an angle of about $24/n$ degrees.

5. Calculation of random orthogonal matrices

Since the elements in successive columns of an orthogonal 3×3 matrix can be regarded as the components of three orthogonal unit vectors, a random orthogonal matrix can be constructed as follows. Choose a random unit vector $\mathbf{x}$ and write its components as the first column. Choose a second random vector $\mathbf{y}'$ which is independent of $\mathbf{x}$. These two vectors define a random plane and in this plane there is a unit vector $\mathbf{y}$ perpendicular to $\mathbf{x}$. The components of $\mathbf{y}$ form the second column while the third column consists of the components of $\mathbf{x} \times \mathbf{y} = \mathbf{z}$ which is normal to the random plane. Thus, the problem is reduced to that of computing the components of a random unit vector.

Let x_1, x_2, x_3 be three independent unit normal deviates with joint probability density

$$(2\pi)^{-\frac{3}{2}} \exp(-\tfrac{1}{2}\Sigma x_r^2).$$

This density is constant on the surface of the sphere $\Sigma x_r^2 = \text{constant}$, so that given the value of $S = \Sigma x_r^2$ the probability that the point (x_1, x_2, x_3) lies in any area of the surface of the sphere is simply proportional to that area. Thus, the direction of the vector $[x_1, x_2, x_3]$ is distributed uniformly and

$$\mathbf{x} = [x_1, x_2, x_3]/S^{\frac{1}{2}}$$

is a random unit vector.

Similarly, we can find another independent random unit vector

$$\mathbf{y}' = [y_1, y_2, y_3]/T^{\frac{1}{2}},$$

where $T = \Sigma y_r^2$. Then, if $P = \mathbf{x}.\mathbf{y}'$,

$$\mathbf{y} = [Sy_1 - Px_1, Sy_2 - Px_2, Sy_3 - Px_3]/[S(ST - P^2)]^{\frac{1}{2}}.$$

Finally, the remaining column of the required matrix follows by computing $\mathbf{z} = \mathbf{x} \times \mathbf{y}$. The values of the random normal deviates were taken from the tables of Mahalanobis,

* This remark is due to Dr H. A. David of the Department of Statistics, University of Melbourne.

Bose, Ray & Banerji (1934) and as a check on the overall accuracy of the calculations the column sums c_1, c_2 and c_3 were formed and it was verified that

$$c_1^2+c_2^2+c_3^2 = 3.$$

The standard deviation of divergences from this equality due to rounding off errors is about 2 units in the last figure retained in the matrix elements.

The distributions of $x_r/S^{\frac{1}{2}}$ and of the column sums c are closely related to the t-distribution and in three dimensions are both distributed uniformly* (Cramer, 1946, pp. 240, 387). Thus, although they are not independent, all nine elements of a random orthogonal matrix are uniformly distributed on the range $(-1, 1)$ while the column sums are uniformly distributed on the range $(-\sqrt{3}, \sqrt{3})$.

The 150 matrices were tested for deviations from these predictions by dividing the range of each variable into 10 equal parts and testing for uniformity of distribution by means of a χ^2-test with 9 degrees of freedom. The greatest value of χ^2 obtained from the elements was 15·7 and for the column sums 22·6, while the corresponding mean values of χ^2 were 9·4 and 16·9. However, after permuting the columns in all the six possible ways,† these maxima dropped to 13·1 and 15·9 respectively while the corresponding means were 9·7 and 12·4. Thus there was then no significant deviation from uniformity of distribution at the 5% level ($\chi^2 = 16{\cdot}9$).

The following four rotation matrices are typical of those computed by the above method:

0·8527	0·4846	0·1953	0·2294	−0·6454	−0·7286
0·2780	−0·7374	0·6155	0·9035	0·4196	−0·0872
0·4423	−0·4705	−0·7636	0·3620	−0·6383	0·6794
0·0443	−0·9973	0·0584	−0·3763	0·5764	0·7254
0·9773	0·0554	0·2045	0·7487	0·6504	−0·1285
−0·2072	0·0480	0·9771	−0·5458	0·4947	−0·6763

The authors wish to thank Dr E. J. Williams for suggesting the method used for calculating the random orthogonal matrices and Dr G. J. Ogilvie and Prof. F. C. Frank for discussion. Mr A. W. Davis checked the counting.

REFERENCES

Cramer, H. (1946). *Methods of Mathematical Statistics.* Princeton University Press.
Jeffreys, H. & Jeffreys, B. S. (1946). *Methods of Mathematical Physics.* Cambridge University Press.
Mahalanobis, P. C., Bose, S. S., Ray, P. R. & Banerji, S. K. (1934). *Sankhya*, **1**, 289.
Ogilvie, G. J. (1952). *J. Inst. Met.* **81**, 491.

* Thus, the cosine of the angle between a fixed direction and a random direction is uniformly distributed on the range $(-1, 1)$. Hence, if ξ the cosine of the co-latitude is chosen at random in the range $(-1, 1)$ and the longitude ϕ at random in the range $(-\pi, \pi)$

$$[(1-\xi^2)^{\frac{1}{2}} \cos\phi,\ (1-\xi^2)^{\frac{1}{2}} \sin\phi,\ \xi],$$

is a random unit vector; the sign of the square root is taken positively and negatively at random. This method of calculation of a random unit vector is suitable for high speed computers.

† This removes any bias due to calculating the successive columns in a definite order.

THE DIFFERENCE BETWEEN CONSECUTIVE MEMBERS OF A SERIES OF RANDOM VARIABLES ARRANGED IN ORDER OF SIZE

By J. H. DARWIN

Applied Mathematics Laboratory, Department of Scientific and Industrial Research, Wellington, New Zealand

1. Introduction and summary

Very commonly in analyses of variance a table of, say, treatment means is set out in order of their size and the experimenter desires to find if there is a significant break between successive treatment means. The theoretical problem of finding the distribution of the difference between the mth and $(m+1)$th in a descending series of N from a known distribution $dF(x)$ readily reduces to the evaluation of an integral. For few distribution functions $F(x)$ can this integral be computed in terms of known tabulated functions. The consequent numerical evaluation is often so tedious that special sorts of comparisons between treatment means have been suggested (e.g. see Tukey (1949), when $F(x)$ is the normal distribution function with an unknown variance).

In this note we find by the saddle-point method an approximate formula for the probability of a greater difference than u between the mth highest and the $(m+1)$th highest of a sample of N drawn from $dF(x)$. This method was first used for approximation on the real axis by Laplace (1820), and, in work similar to ours but on the range of the sample, by Cox (1948). The probability we find when $F(x)$ is the normal integral with known variance, is compared with that obtained by use of the limiting distribution of the mth and $(m+1)$th as deduced by Gumbel (1935) for distributions $dF(x)$ satisfying certain conditions.

Numerical calculation for the 'unstudentized' normal case are compared with previous calculations of the same kind done by Irwin (1925) and the extension to the 'studentized' normal case is given without calculations.

2. The saddle-point approximation

Suppose we have N sample values $x_1, x_2, \ldots, x_N$ arranged in descending order of size; suppose the parent distribution of the x_i is $dF(x)$. Then the probability that $x_m - x_{m+1}$ is greater than u is

$$P_1(u) = \frac{N!}{m!\,(N-m-1)!}\int_{-\infty}^{\infty}(1-F(x+u))^m\,(F(x))^{N-m-1}\,dF(x). \tag{1}$$

For the purpose of using the saddle-point method on this integral we set $m = Np$ and suppose N large. This process may be expected to give reasonable results if p is near $\frac{1}{2}$ and N is large. We shall later use the result for quite small values of N and m. The success of the approximation will then depend, especially when p is near 0 or 1, on the relative values of the first and second approximations, both of which we shall find in special cases.

The integrand is of the form

$$\exp\{N[p\log(1-F(x+u))+(1-p)\log F(x)]\}\,f(x)/F(x) = \exp(NG(x))\,f(x)/F(x), \quad \text{say.} \tag{2}$$

We suppose $F(x)$ has continuous derivatives of the first four orders everywhere in its range. Any stationary value of $G(x)$ is a solution for x of the equation

$$\frac{pf(x+u)}{1-F(x+u)} = \frac{(1-p)f(x)}{F(x)}. \tag{3}$$

Suppose there is a unique root at $\hat{x}$. A useful approximation to it is

$$\hat{x} = x_p + a_1 u + a_2 u^2 + \ldots,$$

where

$$F(x_p) = 1-p,$$

$$a_1 = -(1-p)\,[1+pf'(x_p)/(f(x_p))^2],$$

and $$a_2(f(x_p))^2 + f''(x_p)\,p(1-p)\,(a_1+\tfrac{1}{2}) + f(x_p)f'(x_p)\,(\tfrac{3}{2}a_1^2 + (2-p)\,a_1 + \tfrac{1}{2}(1-p)) = 0.$$

The first approximation to $P(u)$ is then

$$\frac{N!}{m!\,(N-m-1)!} \frac{\sqrt{2\pi}\,(1-F(\hat{x}+u))^m\,(F(\hat{x}))^{N-m-1} f(\hat{x})}{\left[m\left(\frac{f(\hat{x}+u)}{1-F(\hat{x}+u)}\right)^2 + m\left(\frac{f'(\hat{x}+u)}{1-F(\hat{x}+u)}\right) + (N-m)\left(\frac{f(\hat{x})}{F(\hat{x})}\right)^2 - (N-m)\frac{f'(\hat{x})}{F(\hat{x})}\right]^{\frac{1}{2}}}, \tag{4}$$

where $f'(\hat{x})$ and $f'(\hat{x}+u) = \partial f/\partial x$ calculated at $\hat{x}$ and $\hat{x}+u$. Suppose this is written $P_1(u)$. Then the next approximation to $P(u)$ is $P_1(u)\,(1+a/N)$, say, where a is

$$\frac{-5(G'''(\hat{x}))^2}{24(G''(\hat{x}))^3} + \frac{G^{\mathrm{iv}}(\hat{x})}{8(G''(\hat{x}))^2} + \frac{pF(\hat{x})}{2(G''(\hat{x}))^2 f(\hat{x})}(X_p''' X_{1-p}'' - X_{1-p}''' X_p''). \tag{5}$$

In this $X_p^{(i)}$ and $X_{1-p}^{(i)}$ are respectively the coefficients of p and $1-p$ in $G^{(i)}(x)$, the ith partial derivative of $G(x)$ with respect to x which is set equal to $\hat{x}$ after differentiation. For a given F, a is a function of p and u only and can be calculated if F and its derivatives have been tabulated.

3. The normal integral

For applications the most important function $F(x)$ is probably the normal integral

$$F(x) = \frac{1}{\sigma\sqrt{(2\pi)}} \int_{-\infty}^{x} e^{-\frac{1}{2}x^2/\sigma^2}\,dx.$$

We consider first that σ is one. For this distribution $f(x)/F(x)$ is a monotonic decreasing function of x for all real x. This is clearly true for $x \geqslant 0$. For $x < 0$ the derivative of $f(x)/F(x)$ has the sign of

$$-x\int_{-\infty}^{x} (x\,e^{-\frac{1}{2}x^2}/x)\,dx - e^{-\frac{1}{2}x^2} = x\int_{-\infty}^{x} \frac{e^{-\frac{1}{2}x^2}}{x^2}\,dx < 0. \tag{6}$$

Hence as $$f(x+u)/(1-F(x+u)) = f(-(x+u))/F(-(x+u))$$

it follows that $f(x)/F(x)/f(x+u)/(1-F(x+u))$ is a monotonic decreasing function of x. It decreases from infinity to 0 as x increases from minus to plus infinity. Hence there is a unique real root of (3).

By virtue of (6), $G''(\hat{x})$ is less than 0 for all positive u, and the path of the saddle-point method is the real axis from minus to plus infinity. Since $f(x)$ is symmetrical about 0, $\hat{x}(p) + \hat{x}(1-p) = -u$, where the dependence of $\hat{x}$ on p is expressed by writing $\hat{x}$ as $\hat{x}(p)$. It follows easily that $a(p) = a(1-p)$, where $a(p)$ expresses a as a function of p. Some

calculations have been made from the formula (5) for values of p from 0·01 to 0·50 and for values of u which span the 5 % and 1 % significance points for N up to 10 and probably for N up to 50 (to judge from the approximation in § 4); ma/N turns out to vary surprisingly little for a particular value of u and m/N varies from $1/N$ to $\frac{1}{2}$. The calculated values of ma/N are given in Table 1.

Table 1. *Values of* ma/N

u \ p	0·01	0·02	0·05	0·10	0·20	0·30	0·50
0·1	0·0805	—	—	0·0800	—	—	0·1159
0·2	0·0775	—	0·0762	0·0758	0·0773	0·0819	0·1076
0·5	0·0697	—	—	0·0653	0·0646	0·0670	0·0863
1·0	0·0597	0·0575	0·0547	0·0518	0·0489	0·0488	0·0607
2·0	0·0450	0·0425	0·0387	0·0345	0·0297	0·0275	0·0321
3·0	0·0362	0·0330	0·0286	0·0243	0·0193	0·0165	0·0175

We could use either of the equations

$$P_1(u) = 0{\cdot}05 \quad (\text{or } 0{\cdot}01), \tag{7}$$

or

$$P_1(u)\,(1+a/N) = 0{\cdot}05 \quad (\text{or } 0{\cdot}01), \tag{8}$$

to give approximate significance points u. Naturally, one expects (8) to yield values of u closer to the true significance points. The evidence that we have for the accuracy of the points given by (7) and (8) is a comparison of the probabilities $P_1(u)$ and $P_1(u)\,(1+a/N)$ with the probabilities for two distributions worked out by Irwin (1925). These distributions were of the difference between the two largest observations in samples of sizes 3 and 10. The comparison is given in Table 2.

Table 2. *Comparison of approximations with Irwin's exact values*

	u ...	0·1	0·2	0·3	0·4	0·5	0·7	1·0	2·0	3·0
$N=3$	Irwin's values	0·917	0·836	0·760	0·687	—	0·493	0·339	0·069	0·008
	$P_1(u)$	0·8389	0·7699	0·7031	0·6389	—	0·4649	0·3228	0·0672	0·0079
	$P_1(u)\,(1+a/N)$	0·9148	0·8348	0·7584	0·6857	—	0·4928	0·3388	0·0691	0·0081
$N=10$	Irwin's values	0·855	0·727	0·613	—	0·427	—	0·152	0·011	—
	$P_1(u)$	0·7904	0·6734	0·5698	—	0·3994	—	0·1450	0·0110	—
	$P_1(u)\,(1+a/N)$	0·8536	0·7245	0·6108	—	0·4255	—	0·1525	0·0113	—

One might expect that these distributions, corresponding respectively to a low value of N and a low value of p, would provide a severe test of our approximation. The success of $P_1(u)\,(1+a/N)$ over the whole range of u and of $P_1(u)$ for the higher values of u suggests that (8) will give values of u sufficiently correct for practical purposes for N between 3 and 10, and that we will be very close to these values if we use (7). We give in Table 3 the significance points correct to two places for N from 3 to 10, deduced from (8). The figures deduced from (7) are given in brackets.

The solution of (3) necessary to give the tables was made from the *Biometrika* Tables of the normal integral and frequency function, by prospecting till $\hat{x}$ was contained between two neighbouring two-decimal values of the variable. Linear inverse interpolation between these values, at least for $\hat{x}$ up to 2·5, gave a solution $\hat{x}$ which appeared to be correct to 5 decimal places.

Table 3. *Significance points for* $u = x_m - x_{m+1}$

N \ m		1	2	3	4	5
3	1%	2·91 (2·90)				
	5%	2·17 (2·16)				
4	1%	2·60 (2·59)	2·17 (2·17)			
	5%	1·92 (1·90)	1·58 (1·58)			
5	1%	2·42 (2·41)	1·87 (1·86)			
	5%	1·77 (1·76)	1·34 (1·34)			
6	1%	2·30 (2·29)	1·69 (1·69)	1·56 (1·56)		
	5%	1·67 (1·66)	1·21 (1·20)	1·11 (1·11)		
7	1%	2·21 (2·20)	1·58 (1·57)	1·39 (1·39)		
	5%	1·60 (1·59)	1·12 (1·11)	0·98 (0·98)		
8	1%	2·15 (2·13)	1·49 (1·49)	1·28 (1·28)	1·23 (1·22)	
	5%	1·55 (1·53)	1·06 (1·05)	0·90 (0·89)	0·86 (0·85)	
9	1%	2·09 (2·08)	1·43 (1·42)	1·20 (1·20)	1·12 (1·11)	
	5%	1·50 (1·49)	1·01 (1·00)	0·84 (0·83)	0·78 (0·77)	
10	1%	2·04 (2·03)	1·38 (1·37)	1·14 (1·14)	1·04 (1·04)	1·01 (1·01)
	5%	1·46 (1·45)	0·97 (0·96)	0·79 (0·79)	0·72 (0·72)	0·70 (0·70)

In making Table 3 we found two neighbouring two-decimal values of u for which $P_1(u)$ lay just above and just below 0·05 or 0·01. From these we could get the figure in brackets correct to at least 3 decimal places. Where necessary for the rounding off to 2 places, $P_1(u)$ for values of u between these was computed.

Interpolation was made in a graph drawn from the data of Table 1 to give $1+a/N$. Upper and lower estimates were made of the true value, these being taken absurdly far apart to ensure that the true value lay between them. Then it was usually possible to find an unequivocal solution to (8) correct to two decimal places. Any doubtful cases were resolved by the computation of a/N which always lay nearly half-way between the estimates obtained from the graph.

3·1. *The studentized test*

In the application of Table 3 to testing the differences between successive means in an analysis of variance we should use means standardized by division by the known standard deviation of the means. Sometimes previous experience with the material of the experiment will provide a reliable estimate of the standard deviation. More usually, however, the standard deviation will be estimated from some error variance having ν degrees of freedom. For finite but large ν, one may again use the saddle-point method to provide a series in $1/\nu$ giving the probability that $\sqrt{n}\,(x_m - x_{m+1})/s$ is greater than u. In this case x_m is the mth highest of N means, each of n observations, and s^2 is an estimate of the error

variance, with νs^2 having a χ^2 distribution with ν degrees of freedom. Alternatively, one may read this series from Hartley (1944).

We have

$$\text{Prob}\,(\sqrt{n}\,(x_m - x_{m+1}) > us) = P(u) + \left\{\frac{1}{4\nu}(u^2P'' - uP') + \frac{1}{\nu^2}\left[\frac{u^4}{32}P^{\text{iv}} - \frac{u^3}{48}P''' - \frac{u^2}{32}P'' + \frac{u}{32}P'\right] + \ldots\right\}, \tag{9}$$

where $P^{(i)} = (d/du)^i P(u)$ and $P(u)$ is as in (1). The derivatives of $P(u)$ may be obtained from (1) by differentiation under the integral sign. Each ratio $P^{(i)}/P$ may then be approximated by the method of § 2 by a power series in $f(\hat{x}+u)/(1-F(\hat{x}+u))$, where $\hat{x}$ is again the solution of (3) and $P(u)$ is approximated by $P_1(u)$. No calculations have been made for the studentized case, so that the accuracy of this last suggested approximation is unknown.

When there is not enough previous experience for σ to be well enough known to be assumed constant a conservative procedure will be to use, not s, but an upper confidence limit for σ. Thus if we take, say, σ_1 as a 95 % upper confidence limit for σ 19 times out of 20 when we make the statement—the probability of getting a value of u larger than $\sqrt{n}\,(x_m - x_{m+1})/\sigma_1$, is less than (1)—we shall be right. However, this use of Table 3 will not give much discrimination between means for low ν, and its chief advantage over the use of significant differences obtained from the t distribution is that it does produce the correct pattern of the strong dependence of the test values on m.

4. GUMBEL'S ASYMPTOTIC THEORY

An interesting comparison is available with the asymptotic theory of the distribution of the mth value developed by Gumbel (1935). He discusses distributions of what he calls exponential type. For $F(x)$ to be of this type we must have approximately for large x

$$\left.\begin{aligned} -\frac{f(x)}{1-F(x)} &= \frac{f'(x)}{F'(x)}, \\ \text{and} \qquad f^{\nu}(x) &\simeq (-)^{\nu}\frac{(f(x))^{\nu+1}}{(1-F(x))^{\nu}}. \end{aligned}\right\} \tag{10}$$

(10) is satisfied by $1-F(x) = \exp a(x-b)$ $(a<0,\ x>b)$. It is not to be expected that it will be a good approximation for the whole range of x for a distribution as different from the exponential as is the normal. However, for our problem the whole range of x is not so important in the evaluation of (1) as is the range around the expected values of the mth and $(m+1)$th highest values. Gumbel sets $F(u_m) = 1 - m/N$ and expands $F(x)$ about u_m in a Taylor's series, getting, by virtue of (10),

$$F(x) \simeq 1 - \frac{m}{N}\exp\left[-(x-u_m)\,f(u_m)\,N/m\right]. \tag{11}$$

Suppose this is a good enough approximation to $F(x)$ for it to be used in the integration leading to (1). Then the probability that $x_m - x_{m+1}$ is greater than u is approximately

$$\exp\left[-Nuf(u_m)\right]. \tag{12}$$

We give in Table 4 the percentage points calculated from (12) for the normal distribution and the same cases as those of Table 3.

It appears that for the normal distribution (12) is conservative in claiming significance, but is not accurate enough for ordinary use. If $F(x) = 1-\exp(-\alpha x)$, (12) is exact.

Table 4. *Significance points for $x_m - x_{m+1}$ given by Gumbel's theory*

N \ m		1	2	3	4	5
3	1%	4·22				
	5%	2·75				
4	1%	3·62	2·89			
	5%	2·36	1·88			
5	1%	3·29	2·38			
	5%	2·14	1·55			
6	1%	3·07	2·11	1·92		
	5%	2·00	1·37	1·25		
7	1%	2·92	1·94	1·68		
	5%	1·90	1·26	1·09		
8	1%	2·80	1·81	1·52	1·44	
	5%	1·82	1·18	0·99	0·94	
9	1%	2·70	1·72	1·41	1·30	
	5%	1·76	1·12	0·92	0·84	
10	1%	2·62	1·64	1·32	1·19	1·15
	5%	1·71	1·07	0·86	0·78	0·75

5. Other distribution functions

The exponential distribution provides a comparison of the theoretical probability with the known probability (12), which is $\exp(-\alpha mu)$. The approximation by $P_1(u)$ is

$$\binom{N}{m}\sqrt{(2\pi N)}\,p^{m+\frac{1}{2}}(1-p)^{N-m+\frac{1}{2}}\exp(-\alpha mu). \tag{13}$$

This coincides with the correct value if the first term of Stirling's approximation to the gamma function is used for the three terms of the binomial coefficient. The second approximation is (13) multiplied by

$$1+\frac{a(p)}{N} = 1+\frac{p^2-p+1}{12Np(1-p)}, \tag{14}$$

balancing the second term in Stirling's approximation.

Again $a(p) = a(1-p)$, so that the correction is still symmetrical about $p = \frac{1}{2}$, although the main term $\exp(-\alpha mu)$ is not symmetrical about $m = \frac{1}{2}N$. $a(p)/N$ is still small, achieving its highest value $\frac{1}{12}$ for $p = 1/N$ and N large. (A similar analysis can be done when $F(x) = \exp(-\exp(-\alpha x))$.) A test that is independent of the value of α is possible if an independent estimate of $1/\alpha$ is available from, say, the average of r independent variables drawn from the distribution. The sum of r such values, X, say, when multiplied by α has the distribution $(1/\Gamma(r))\exp(-x)\,x^{r-1}\,dx$ and the probability that $(x_m - x_{m+1})/X$ is greater than u is

$$(1+mu)^{-r}. \tag{15}$$

More usually no such independent sample will be available, and it is perhaps more appropriate then to test the ratio x_m/x_{m+1} for $\alpha = 1$. Instead of equation (3), we have the equation

$$\frac{pcf(cx)}{1-F(cx)} = \frac{(1-p)f(x)}{F(x)}, \tag{16}$$

when we are trying to find the probability that x_m is greater than cx_{m+1}. This is satisfied by $\exp(-x) = pc/(pc+1-p)$. The first approximation to the probability is

$$\binom{N}{m}\sqrt{(2\pi N)}\,(pc)^{cm+\frac{1}{2}}(1-p)^{N-m+\frac{1}{2}}(pc+1-p)^{-N+m-cm-\frac{1}{2}}, \tag{17}$$

and the correction factor is $$1+\frac{(1-p)^2+pc(pc-1+p)}{12Ncp(1-p)(pc+1-p)}, \tag{18}$$

where $a(p,c) = ca(1-p, 1/c)$. The biggest value of a/N is again $\frac{1}{12}$. The true value for $P(u)$ is

$$\binom{N}{m}\Big/\binom{N+mc-m}{mc}, \tag{19}$$

and the correction factor balances the second term in Stirling's approximation to the terms of the denominator of (19).

5·2. *The rectangular distribution*

Another distribution that provides a check on the accuracy of $P_1(u)$ and $P_1(u)(1+a/N)$ is the rectangular distribution, $F(x) = x$ in $(0, 1)$. Then $P_1(u)$ is

$$\binom{N}{m}\sqrt{(2\pi N)}\,p^{m+\frac{1}{2}}(1-p)^{N-m+\frac{1}{2}}(1-u)^N, \tag{20}$$

and as in (14)
$$\frac{a(p)}{N} = \frac{1-p+p^2}{12Np(1-p)}.$$

The true value of $P(u)$ is $(1-u)^N$. Thus for both exponential and rectangular distributions and for the range of u investigated for the normal distribution the correction factor is less than $\frac{1}{12}$.

6. RENORMALIZATION

As suggested by the referee, we can compare the above results with those obtained by renormalizing $P_1(u)$ (see Cox, 1948; Daniels, 1956). We renormalize $P_1(u)$ by dividing by $P_1(0)$. When $u = 0$, $F(\hat{x}) = 1-p$, and, as u tends to 0, $a(u)/N$ tends to

$$(1-p+p^2)/\{12Np(1-p)\}$$

which agrees with $1/P_1(0)$ to order $1/N$.

$$P_1(0) = \binom{N}{m}\sqrt{(2\pi N)}\,p^{m+\frac{1}{2}}(1-p)^{N-m+\frac{1}{2}}. \tag{21}$$

In general, $a(u)/N$ seems to decrease as u increases. Hence use of $P_1(u)/P_1(0)$ instead of $P_1(u)(1+a/N)$ will tend to overestimate the probability of a bigger difference than u. This difference between the two, however, is likely to be numerically more serious for $P_1(u)$ near $\frac{1}{2}$ than for $P_1(u)$ near 0, since when $P_1(u)$ tends to zero the difference also tends to zero. This is illustrated by the values of $P_1(u)/P_1(0)$ for the same values of u as in Table 2:

$N = 3$:	0·9223	0·8464	0·7730	0·7024	0·5111	03549	0·0739	0·0087
$N = 10$:	0·8579	0·7309	0·6185	0·4335	0·1574	0·0119		

The pattern of the difference for $m \neq 1$ can be gauged from the sizes of the factors $1+a/N$ and $1/P_1(0)$ for $N = 20$. These are:

$1+a/N$ from $u = 0{\cdot}2$ to $u = 3{\cdot}0$ decreases	$1/P_0$
for $m = 1$ from 1·0762 to 1·0286	1·0847
$m = 2$ from 1·0379 to 1·0122	1·0427
$m = 4$ from 1·0193 to 1·0048	1·0221
$m = 6$ from 1·0137 to 1·0028	1·0158
$m = 10$ from 1·0108 to 1·0018	1·0126

The biggest differences between the factors occur for low values of m and high values of u. For higher u, near the significance points, there is little to choose between $P_1(u)$, $P_1(u)(1+a/N)$ and $P_1(u)/P_1(0)$, since the difference between any pair is, for example, of order $(0{\cdot}05)(1+\frac{1}{12})$ for the 5 % point; but for low values of u the second two are appreciably more accurate than the first.

My thanks are due to Miss Mary Chung for help with the calculations.

REFERENCES

Cox, D. R. (1948). A note on the asymptotic distribution of range. *Biometrika*, **35**, 310–15.

Daniels, H. E. (1956). The approximate distribution of serial correlation coefficients. *Biometrika*, **43**, 169–85.

Gumbel, E. J. (1935). Les valeurs extrêmes des distributions statistiques. *Ann. Inst. Henri Poincaré*, **5**, 115–58.

Hartley, H. O. (1943). Studentization or the elimination of the standard deviation of the parent population from the random sample-distribution of statistics. *Biometrika*, **33**, 173–80.

Irwin, J. O. (1925). On a criterion for the rejection of outlying observations. *Biometrika*, **17**, 238–50.

Laplace, P. S. (1820). *Théorie Analytique des Probabilités*, **1**, part 2, chap. 1.

Tukey, J. W. (1949). Comparing individual means in the analysis of variance. *Biometrics*, **5**, 99–114.

RELATION BETWEEN THE DISTRIBUTIONS OF NON-CENTRAL t AND OF A TRANSFORMED CORRELATION COEFFICIENT

By B. I. HARLEY

University College, London

1. If r is the correlation coefficient in a sample of size n drawn randomly from a normal bivariate population with zero correlation, then the quantity $(n-2)^{\frac{1}{2}} r/(1-r^2)^{\frac{1}{2}}$ is known to be distributed exactly as Student's t with $(n-2)$ degrees of freedom.

By a comparison of moments we shall show that if the population correlation, ρ, is not zero, then

$$v = \frac{r}{(1-r^2)^{\frac{1}{2}}}(n-2)^{\frac{1}{2}} g\{\rho\},$$

where $g\{\rho\}$, an appropriately chosen function of ρ, is distributed approximately as non-central t. We define the distribution of non-central t having f degrees of freedom as that of

$$t = \frac{z+\delta}{w^{\frac{1}{2}}},$$

where z is a unit normal variable having zero expectation, w is an independent variable distributed as χ^2/f with f degrees of freedom and δ is the non-central parameter. We shall suggest a form for $g\{\rho\}$ and determine appropriate relations between n and ρ on the one hand and f and δ on the other. Finally, some numerical comparisons will be given and a method indicated whereby the approximate equivalence of the distributions may be used to supplement the tables of the non-central t distribution given by Johnson & Welch (1939).

2. Our first objective will be to obtain the moments of $r/(1-r^2)^{\frac{1}{2}}$ when $\rho \neq 0$. The probability distribution of r is known to be

$$p(r \mid n\rho) = \frac{(1-\rho^2)^{\frac{1}{2}(n-1)}}{\pi(n-3)!}(1-r^2)^{\frac{1}{2}(n-4)} \frac{d^{n-2}}{d(r\rho)^{n-2}}\left\{\frac{\cos^{-1}(-r\rho)}{(1-r^2\rho^2)^{\frac{1}{2}}}\right\}. \tag{1}$$

By integrating over the range of values of r and denoting $\dfrac{d^{n-2}}{d(r\rho)^{n-2}}\left\{\dfrac{\cos^{-1}(-r\rho)}{(1-r^2\rho^2)^{\frac{1}{2}}}\right\}$ by $f^{n-2}(r\rho)$ we find that

$$\int_{-1}^{+1} (1-r^2)^{\frac{1}{2}(n-4)} f^{n-2}(r\rho)\, dr = \frac{\pi(n-3)!}{(1-\rho^2)^{\frac{1}{2}(n-1)}} = C(n,\rho).$$

Differentiating with respect to ρ gives

$$\int_{-1}^{+1} (1-r^2)^{\frac{1}{2}(n-4)} r \frac{\partial}{\partial(r\rho)}[f^{n-2}(r\rho)]\, dr = \frac{(n-1)\pi(n-3)!\rho}{(1-\rho^2)^{\frac{1}{2}(n+1)}} = C^1(n,\rho),$$

where
$$C^m(n,\rho) = \frac{\partial^m}{\partial\rho^m} C(n,\rho) \quad \text{for} \quad m = 1, 2, \ldots.$$

Hence
$$\int_{-1}^{+1} \frac{r}{(1-r^2)^{\frac{1}{2}}}(1-r^2)^{\frac{1}{2}(n-3)} f^{n-1}(r\rho)\, dr = C^1(n,\rho). \tag{2}$$

If we denote the expectation in a sample of size k by $\mathscr{E}_k$, where k can take integral values, then

$$\mathscr{E}_{n+1}\left(\frac{r}{(1-r^2)^{\frac{1}{2}}}\right) = \int_{-1}^{+1} \frac{r}{(1-r^2)^{\frac{1}{2}}} \frac{(1-r^2)^{\frac{1}{2}(n-3)}}{C(n+1,\rho)} f^{n-1}(r\rho)\,dr$$

$$= \frac{C^1(n,\rho)}{C(n+1,\rho)}.$$

If $(n+1)$ is replaced by n, then

$$\mathscr{E}_n\left(\frac{r}{(1-r^2)^{\frac{1}{2}}}\right) = \frac{C^1(n-1,\rho)}{C(n,\rho)}. \tag{3}$$

To obtain the second moment we consider

$$C^2(n-2,\rho) = \frac{\partial}{\partial\rho} C^1(n-2,\rho) = \int_{-1}^{+1} \frac{\partial}{\partial\rho} \frac{r}{(1-r^2)^{\frac{1}{2}}} (1-r^2)^{\frac{1}{2}(n-5)} f^{n-3}(r\rho)\,dr$$

$$= \int_{-1}^{+1} \frac{r^2}{(1-r^2)} (1-r^2)^{\frac{1}{2}(n-4)} f^{n-2}(r\rho)\,dr$$

$$= C(n,\rho)\,\mathscr{E}_n\left(\frac{r^2}{1-r^2}\right).$$

Hence
$$\mathscr{E}_n\left(\frac{r^2}{1-r^2}\right) = \frac{C^2(n-2,\rho)}{C(n,\rho)}. \tag{4}$$

In general we have
$$\mathscr{E}_n\left(\frac{r}{(1-r^2)^{\frac{1}{2}}}\right)^m = \frac{C^m(n-m,\rho)}{C(n,\rho)}, \tag{5}$$

since differentiating with respect to ρ results in effect to multiplying by r, and increasing the power of the differential by one.

From equation (5) the first four moments about the origin of

$$v = (n-2)^{\frac{1}{2}} \frac{r}{(1-r^2)^{\frac{1}{2}}} g\{\rho\}$$

are found to be

$$\mu_1'(v) = \mathscr{E}_n(v) = \frac{(n-2)^{\frac{1}{2}}}{(n-3)} \frac{\rho}{(1-\rho^2)^{\frac{1}{2}}} g\{\rho\}, \tag{6}$$

$$\mu_2'(v) = \mathscr{E}_n(v^2) = \frac{(n-2)}{(n-4)}\left(1 + \frac{(n-1)\rho^2}{(1-\rho^2)}\right)[g\{\rho\}]^2, \tag{7}$$

$$\mu_3'(v) = \mathscr{E}_n(v^3) = \frac{(n-2)^{\frac{3}{2}}}{(n-3)(n-5)} \frac{\rho}{(1-\rho^2)^{\frac{1}{2}}}\left(3 + \frac{n\rho^2}{(1-\rho^2)}\right)[g\{\rho\}]^3, \tag{8}$$

$$\mu_4'(v) = \mathscr{E}_n(v^4) = \frac{(n-2)^2}{(n-4)(n-6)}\left(3 + \frac{6(n-1)\rho^2}{(1-\rho^2)} + \frac{(n^2-1)\rho^4}{(1-\rho^2)^2}\right)[g\{\rho\}]^4. \tag{9}$$

From these equations we obtained the moments about the mean and the beta-ratios, but since they are not necessary for the subsequent work, they are not reproduced here owing to their length.

3. Johnson & Welch gave the first three moments of the non-central t distribution, and since w and z are independent it is a straightforward matter to find the fourth moment and the beta-ratios.

We find that, for the moments about $t = 0$,

$$\mu_1'(t) = \frac{f^{\frac{1}{2}}}{2^{\frac{1}{2}}} \frac{\Gamma(\frac{1}{2}f - \frac{1}{2})}{\Gamma(\frac{1}{2}f)} \delta, \tag{10}$$

$$\mu_2'(t) = \frac{f}{f-2}(1+\delta^2), \tag{11}$$

$$\mu_3'(t) = \frac{f^{\frac{3}{2}}}{2^{\frac{3}{2}}} \frac{\Gamma(\frac{1}{2}f - \frac{3}{2})}{\Gamma(\frac{1}{2}f)} \delta(3+\delta^2), \tag{12}$$

$$\mu_4'(t) = \frac{f^2}{(f-2)(f-4)}(3+6\delta^2+\delta^4), \tag{13}$$

and from these equations we obtained the moments about the mean and the beta-ratios.

4. To bring the distributions of $r/(1-r^2)^{\frac{1}{2}}$ and of non-central t into correspondence we have to specify a form for the function $g\{\rho\}$ and to relate n and ρ to f and δ. It is clear that the link could be made in a number of ways. After various methods had been examined that described below was chosen, partly because it led to simple relations which did not involve any approximation for the Gamma function and also because it appeared to provide satisfactory results in the cases where numerical comparisons were made.

From equations (10) and (12) we have

$$\frac{\mu_3'(t)}{\mu_1'(t)} = \frac{f}{(f-3)}(3+\delta^2), \tag{14}$$

and from equations (6) and (8) it follows that

$$\frac{\mu_3'(v)}{\mu_1'(v)} = [g\{\rho\}]^2 \frac{(n-2)}{(n-5)}\left(3 + \frac{n\rho^2}{(1-\rho^2)}\right). \tag{15}$$

Equating these moment ratios gives

$$[g\{\rho\}]^2 \frac{(n-2)}{(n-5)}\left(3 + \frac{n\rho^2}{(1-\rho^2)}\right) = \frac{f}{(f-3)}(3+\delta^2). \tag{16}$$

Next equating the values of $\mu_2'(v)$ and $\mu_2'(t)$ given in equations (7) and (11) we have

$$[g\{\rho\}]^2 \frac{(n-2)}{(n-4)}\left(1 + \frac{(n-1)\rho^2}{(1-\rho^2)}\right) = \frac{f}{(f-2)}(1+\delta^2). \tag{17}$$

If we take $f = n-2$, a result which gives exact correspondence when $\rho = 0$, then (16) and (17) are satisfied when

$$\delta = \left(\frac{(2n-3)\rho^2}{(2-\rho^2)}\right)^{\frac{1}{2}} \quad \text{and} \quad g\{\rho\} = \left(\frac{2(1-\rho^2)}{(2-\rho^2)}\right)^{\frac{1}{2}}. \tag{18}$$

Apart from the second moments about zero, the moments of the two corresponding distributions will not agree exactly, but some results given in Table 1 suggest that the representation of one distribution by the other may nevertheless be quite good.

Table 1. *Comparison of the values of the first two moments and $\sqrt{\beta_1}$ and β_2 of v and t*

ρ	n	δ	$\mu_1'(v)$	$\mu_1'(t)$	$\mu_2(v)$	$\mu_2(t)$	$\sqrt{\beta_1}\,(v)$	$\sqrt{\beta_1}\,(t)$	$\beta_2(v)$	$\beta_2(t)$
0·2	15	0·742	0·7891	0·7889	1·2103	1·2107	0·2150	0·2139	3·7386	3·7352
0·2	27	1·020	1·0522	1·0521	1·1111	1·1113	0·1361	0·1356	3·3121	3·3115
0·6	15	2·435	2·588	2·587	1·489	1·493	0·631	0·615	4·260	4·236
0·6	27	3·346	3·451	3·450	1·346	1·353	0·404	0·369	3·556	3·627

5. Johnson & Welch considered the case of calculating a value t_ϵ such that $P\{t > t_\epsilon\} = \epsilon$, where δ and f are assumed known, and they gave tables of a function $\lambda(f, t_\epsilon, \epsilon)$ for various probability levels ϵ. By an inverse use of these tables and an iterative process, a value of t_ϵ can be calculated when f and δ are known, for the various probability levels ϵ given. If t_ϵ is required for other probability levels an estimate of t_ϵ could be made using the normal equivalent deviate as an estimate of λ, but this does not give a very good approximation in most cases.

The probability integral of r in samples of size n from a normal bivariate population has been calculated by F. N. David (1938) for values of $n = 3\,(1)\,25, 50, 100, 200, 400$ and $\rho = 0{\cdot}0\,(0{\cdot}1)\,0{\cdot}9$, and thus by using the approximate connexion between r and t given in § 4 we can obtain values of t_ϵ, given δ and f for any probability level and not only those given in Johnson & Welch's tables.

The method of calculation of t_ϵ, given δ, f and ϵ, is as follows:

(i) Using the relationship between δ and ρ given in equation (18) we have that

$$\rho = \left(\frac{2\delta^2}{2n-3+\delta^2}\right)^{\frac{1}{2}},$$

where $n = f+2$, and where from a consideration of the values of $\mu_1'(v)$ and $\mu_1'(t)$ we note that ρ and δ have the same sign.

(ii) Using this value of ρ, and taking $n = f+2$, a value r_ϵ can be found from David's tables such that $P\{r > r_\epsilon\} = \epsilon$.

(iii) Finally, t_ϵ is calculated from the relationship

$$t_\epsilon = \left(\frac{2f(1-\rho^2)}{2-\rho^2}\right)^{\frac{1}{2}} \frac{r_\epsilon}{(1-r_\epsilon^2)^{\frac{1}{2}}}.$$

To illustrate the accuracy of the approximation for given values of δ, f and ϵ, values of t_ϵ were found by means of the r approximation and compared with the exact values found by using Johnson & Welch's Table IV. For comparative purposes the values of t_ϵ using Johnson & Welch's first approximation, in which the equivalent normal deviate was taken as an estimate of λ, were also calculated. To simplify the entry into David's tables δ was chosen in such a way that 10ρ was integral, but, of course, this is not a necessary condition. From the nature of Johnson & Welch's Table IV it would seem reasonable to expect in most cases an accurate estimate of t_ϵ to three significant figures only. From the results given in Table 2, where the number of figures given is determined by the number given for λ by Johnson & Welch, it is seen that the values of t_ϵ calculated by using Johnson & Welch's

Table 2. *Comparison of the values of t_ϵ calculated by three different methods at various probability levels*

f	δ	ρ	ϵ	Values of t_ϵ		
				Exact	From τ approximation	From Johnson & Welch's first approximation
7	0·553	0·2	0·4	0·843	0·843	0·813
			0·2	1·512	1·513	1·457
			0·05	2·621	2·623	2·542
			0·01	3·90	3·90	3·93
	1·815	0·6	0·4	2·190	2·188	2·105
			0·2	2·992	2·992	2·876
			0·05	4·434	4·436	4·331
			0·01	6·20	6·20	6·45
	2·657	0·8	0·4	3·105	3·100	2·981
			0·2	4·030	4·026	3·867
			0·05	5·739	5·741	5·628
			0·01	7·87	7·89	8·34
	3·195	0·9	0·4	3·699	3·688	3·544
			0·2	4·708	4·701	4·512
			0·05	6·603	6·608	6·488
			0·01	8·99	9·01	9·60
16	0·821	0·2	0·4	1·096	1·096	1·079
			0·2	1·727	1·727	1·699
			0·05	2·674	2·674	2·635
			0·01	3·60	3·60	3·57
	2·691	0·6	0·4	3·029	3·027	2·978
			0·2	3·760	3·758	3·697
			0·05	4·920	4·922	4·860
			0·01	6·12	6·12	6·12
	3·941	0·8	0·4	4·334	4·330	4·258
			0·2	5·159	5·156	5·071
			0·05	6·502	6·509	6·431
			0·01	7·91	7·94	7·95
	4·739	0·9	0·4	5·173	5·165	5·080
			0·2	6·068	6·063	5·962
			0·05	7·538	7·546	7·462
			0·01	9·10	9·12	9·17

Table IV agree with those calculated using the r approximation, to this accuracy at least, in all cases where $\rho = 0{\cdot}2$ and $0{\cdot}6$. For $\rho = 0{\cdot}8$ and $0{\cdot}9$ the agreement is not quite as good, but for $\epsilon = 0{\cdot}4$, $0{\cdot}2$ and $0{\cdot}05$ there is still only a difference of at most one unit in the third significant figure between the exact value and that from the r approximation. Since Johnson & Welch give fewer decimal places for $\epsilon = 0{\cdot}01$, one would expect that for large values of δ the third significant figure may be in error by one or two units which may account for the seemingly larger discrepancy between the first two methods.

In most cases the method using the probability integral appears to be more accurate than Johnson & Welch's first approximation given in the last column of Table 2. Certainly these results suggest that the probability integral of r could be used to obtain reasonably accurate values of t_ϵ for those values of ϵ not given in Johnson & Welch's Table IV.

REFERENCES

David, F. N. (1938). *Tables of the Ordinates and Probability Integral of the Distribution of the Correlation Coefficient in Small Samples.* Cambridge University Press.

Johnson, N. L. & Welch, B. L. (1939). *Biometrika*, **31**, 362.

ON THE SOLUTION OF ESTIMATING EQUATIONS FOR TRUNCATED AND CENSORED SAMPLES FROM NORMAL POPULATIONS*

By A. CLIFFORD COHEN, Jr.

The University of Georgia

1. Introduction and summary

To calculate maximum-likelihood estimates of the mean and standard deviation of a normally distributed population from doubly truncated or from doubly censored samples, it is necessary to solve simultaneously a pair of rather complex non-linear estimating equations. Although solutions can be approximated by straightforward iterative procedures, the calculations often become tedious and time-consuming. This paper is concerned with reducing computational labour required in obtaining these solutions. For use in the doubly truncated case, a chart which permits direct reading of the standardized terminals with an accuracy of from three to five units in the second decimal is included. For use in both the doubly truncated and doubly censored cases, an iterative procedure is devised, which, for a specified degree of accuracy, appears to require less computational effort than similar procedures previously proposed. For use in singly censored cases, where only one estimating equation is involved, a chart which permits direct reading of estimates of the standardized terminal with a degree of accuracy comparable to that possible in the doubly truncated case is presented. Since tables recently published by Cohen & Woodward (1953) reduce the calculation of estimates from singly truncated normal samples to the simple task of interpolating between table entries, that case is considered here only by reference.

2. Graphic solution of estimating equations for doubly truncated samples

We let x designate a normally distributed random variable with probability density function

$$f(x) = (\sigma\sqrt{(2\pi)})^{-1}\exp[-(x-\mu)^2/(2\sigma^2)] \quad (-\infty < x < \infty). \tag{1}$$

For a doubly truncated sample consisting of n observations from this population, each of which is subject to the restriction $x_0 \leqslant x \leqslant x_0 + w$, where the sample terminals, x_0 and $x_0 + w$, are fixed, the logarithm of the likelihood function is

$$L = -n\ln[I(\xi_1) - I(\xi_2)] - n\ln\sigma - \sum_{i=1}^{n}(x_i-\mu)^2/(2\sigma^2) + \text{const.}, \tag{2}$$

where

$$I(\xi) = \int_{\xi}^{\infty}\phi(t)\,dt, \quad \phi(t) = (\sqrt{(2\pi)})^{-1}\exp -\tfrac{1}{2}t^2, \tag{3}$$

and

$$\xi_1 = (x_0-\mu)/\sigma, \quad \xi_2 = (x_0+w-\mu)/\sigma. \tag{4}$$

* Sponsored by the Office of Ordnance Research, U.S. Army.

Maximum-likelihood estimating equations obtained by equating to zero, the partial derivatives of L with respect to μ and σ respectively were given by Cohen (1950) as

$$\left.\begin{aligned}\sigma[\bar{Z}_1-\bar{Z}_2-\xi_1]-\nu_1=0,\\ \sigma^2[1-\xi_1(\bar{Z}_1-\bar{Z}_2-\xi_1)-\bar{Z}_2 w/\sigma]-\nu_2=0,\end{aligned}\right\} \tag{5}$$

where
$$\bar{Z}_1=\phi(\xi_1)/[I(\xi_1)-I(\xi_2)], \quad \bar{Z}_2=\phi(\xi_2)/[I(\xi_1)-I(\xi_2)], \tag{6}$$

$$\nu_k=\sum_{i=1}^{n}(x_i-x_0)^k/n. \tag{7}$$

No information is assumed to be available about observations which might have been eliminated as a consequence of the restrictions on observation of x.

A procedure based on two-way interpolation was employed by Cohen (1950) to solve equations (5) simultaneously for maximum-likelihood estimates, $\hat{\sigma}$ and $\hat{\xi}_1$. With these values determined, $\hat{\mu}$ follows from (4) as

$$\hat{\mu}=x_0-\hat{\sigma}\hat{\xi}_1. \tag{8}$$

Throughout this paper, the symbol ($\hat{}$) serves to distinguish maximum-likelihood estimates from parameters estimated.

Expressing σ as
$$\sigma=w/(\xi_2-\xi_1), \tag{9}$$

a result which follows from (4), further reduction of estimating equations (5) yields

$$\left.\begin{aligned}[\bar{Z}_1-\bar{Z}_2-\xi_1]/(\xi_2-\xi_1)-\nu_1/w=0,\\ [1+\xi_1\bar{Z}_1-\xi_2\bar{Z}_2-(\bar{Z}_1-\bar{Z}_2)^2]/(\xi_2-\xi_1)^2-s^2/w^2=0,\end{aligned}\right\} \tag{10}$$

where s^2 is the truncated sample variance ($=\nu_2-\nu_1^2$). The two equations of (10) are thus of the form $F_i(\xi_1,\xi_2)=H_i(\xi_1,\xi_2)-K_i=0$ $(i=1,2)$, where for a given sample the K_i are constants.

The above form for the second equation of (10) was suggested by Mr George W. Thomson, who, together with Friedman and Garelis (1954), tabulated $H_1(\xi_1,\xi_2)$ and $H_2(\xi_1,\xi_2)$ at intervals of 0·5 for the two arguments. As a means of circumventing interpolation difficulties imposed by the large tabular intervals and in order to further facilitate the solution of (10), the two families of curves, $F_1(\xi_1,\xi_2)=0$ and $F_2(\xi_1,\xi_2)=0$, were plotted for selected values of ν_1/w and s^2/w^2, and are presented here as Fig. 1.* Co-ordinates for points on these curves were read from large-scale graphs of $H_i(\xi_1,\xi_2)$ carefully plotted from the Thomson, Friedman & Garelis (1954) tables as functions of ξ_1 for the family of ξ_2 values.

The fact that H_1 and H_2 obey the relations

$$\left.\begin{aligned}H_1(\xi_1,\xi_2)=[\bar{Z}_1-\bar{Z}_2-\xi_1]/(\xi_2-\xi_1)=1-H_1(-\xi_2,-\xi_1),\\ H_2(\xi_1,\xi_2)=[1+\xi_1\bar{Z}_1-\xi_2\bar{Z}_2-(\bar{Z}_1-\bar{Z}_2)^2]/(\xi_2-\xi_1)^2=H_2(-\xi_2,-\xi_1),\end{aligned}\right\} \tag{11}$$

permitted one-half the points required in plotting Fig. 1 to be obtained by reflexion. From the first of the above relations, it follows that the graph of $H_1(\xi_1,\xi_2)-K=0$ is the reflexion of the graph of $H_1(\xi_1,\xi_2)-(1-K)=0$ about the line $\xi_1+\xi_2=0$, while from the second it follows that $H_2(\xi_1,\xi_2)$ is symmetric about this line.

* The assistance of Miss Lola Kiser and Mr Robert Lathrop in the preparation of this chart is gratefully acknowledged. Copies of the original 8-figure tables are available and can be obtained on request from Mr G. W. Thomson, Ethyl Corporation, 1600 West Eight Mile Road, Detroit 20, Michigan, U.S.A.

With ν_1/w and s^2/w^2 computed for a given sample, the intersection of corresponding curves of $F_1(\xi_1, \xi_2) = 0$ and $F_2(\xi_1, \xi_2) = 0$ is located on the chart of Fig. 1. Co-ordinates of this point, which can be read to within three to five units in the second decimal, are the required

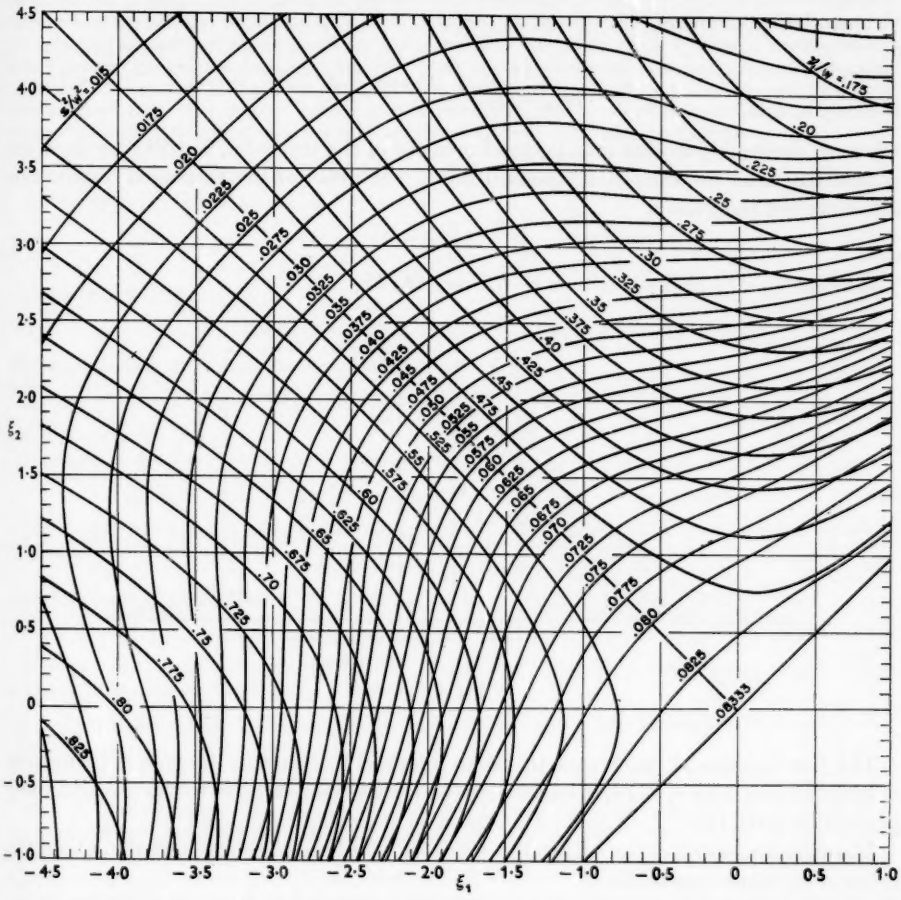


Fig. 1. Estimation curves for doubly truncated samples. (1) Locate curve corresponding to sample value of ν_1/w. (2) Follow curve thus located to its intersection with curve which corresponds to sample value of s^2/w^2. (3) Co-ordinates of intersection determined above, which may be read on scales along base and left edge of chart, are the required values of $\hat{\xi}_1$ and $\hat{\xi}_2$.

values of $\hat{\xi}_1$ and $\hat{\xi}_2$. This degree of accuracy is adequate for many practical applications, but when more precise results are necessary, values thus read serve as first approximations to be improved through iteration.

With $\hat{\xi}_1$ and $\hat{\xi}_2$ determined, estimates of the mean and standard deviation follow from (4) as

$$\hat{\sigma} = w/(\hat{\xi}_2 - \hat{\xi}_1), \quad \hat{\mu} = x_0 - \hat{\sigma}\hat{\xi}_1. \tag{12}$$

Although estimating equations (10) were derived by the method of maximum likelihood, these same equations also result from equating the first and second truncated sample moments to corresponding population moments. In standardized population (complete distribution) units, the mean of the population central block (truncated distribution) from the population mean may be written as

$$\bar{\alpha}_1 = \frac{1}{I(\xi_2) - I(\xi_1)} \int_{\xi_1}^{\xi_2} t\phi(t)\, dt = \bar{Z}_1 - \bar{Z}_2,$$

where $\bar{\alpha}_k$ designates the kth standardized moment of the truncated distribution from the population mean in population standard units. The mean of the truncated distribution about the left terminus is

$$\bar{Z}_1 - \bar{Z}_2 - \xi_1.$$

The foregoing may be made clearer by reference to Fig. 2.

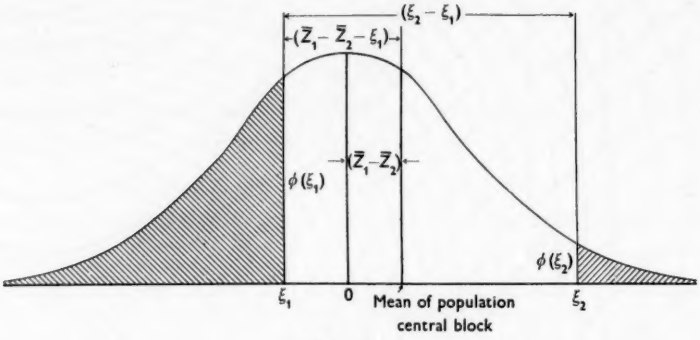


Fig. 2

The first equation of (10) is then the result obtained by equating the ratio of (i) distance of sample mean from x_0 to (ii) distance between truncation points, w, to the corresponding population ratio, $(\bar{Z}_1 - \bar{Z}_2 - \xi_1)/(\xi_2 - \xi_1)$, otherwise denoted as $H_1(\xi_1, \xi_2)$.

The second moment of the central block in a standardized normal population about the mean of the whole population is

$$\bar{\alpha}_2 = \frac{1}{I(\xi_2) - I(\xi_1)} \int_{\xi_1}^{\xi_2} t^2\phi(t)\, dt = 1 + \xi_1\bar{Z}_1 - \xi_2\bar{Z}_2.$$

The variance, $\bar{\sigma}^2$, of the central block of the population is therefore

$$\bar{\sigma}^2 = \bar{\alpha}_2 - \bar{\alpha}_1^2 = 1 + \xi_1\bar{Z}_1 - \xi_2\bar{Z}_2 - (\bar{Z}_1 - \bar{Z}_2)^2.$$

The second equation of (10) then results from equating the ratio of (i) sample variance to (ii) the square of the distance between truncation points, w^2, to the corresponding population ratio $[1 + \xi_1\bar{Z}_1 - \xi_2\bar{Z}_2 - (\bar{Z}_1 - \bar{Z}_2)^2]/(\xi_2 - \xi_1)^2$, otherwise denoted as $H_2(\xi_1, \xi_2)$.

3. ITERATIVE SOLUTIONS

When solutions to the estimating equations for doubly truncated samples obtained by reading coordinates from Fig. 1 as described in § 2 are not of the required accuracy, the standard maximum-likelihood iterative procedure (Newton's method) may be employed to compute corrections to these initial values. However, since the second partial derivatives of the likelihood function must be evaluated, each cycle of the iterative process is likely to prove tedious and laborious. Furthermore, as a consequence of neglecting second and higher powers of the corrections, Newton's method tends to produce rather slowly converging iterants during the first few cycles unless initial approximations are in a close neighbourhood of the solution. This difficulty has been recognized and discussed, for example, by Norton (1956). In an effort to overcome these objections, the method of successive substitutions as described by Scarborough (1930, pp. 191–5) has been employed to develop iterants of the form

$$\xi_1 = f(\xi_1, \xi_2), \quad \xi_2 = g(\xi_1, \xi_2). \tag{13}$$

On clearing of fractions, equations (10) become

$$\left.\begin{aligned} [\overline{Z}_1 - \overline{Z}_2 - \xi_1] &= (\xi_2 - \xi_1)\,\nu_1/w, \\ [1 + \xi_1\overline{Z}_1 - \xi_2\overline{Z}_2 - (\overline{Z}_1 - \overline{Z}_2)^2] &= (\xi_2 - \xi_1)^2\,(s/w)^2. \end{aligned}\right\} \tag{14}$$

Using the identity

$$[1 + \xi_1\overline{Z}_1 - \xi_2\overline{Z}_2 - (\overline{Z}_1 - \overline{Z}_2)^2] \equiv [1 - (\overline{Z}_1 - \overline{Z}_2 - \xi_1)\,(\overline{Z}_1 - \overline{Z}_2) - (\xi_2 - \xi_1)\,\overline{Z}_2],$$

the two equations of (14) may be combined to give

$$(s/w)^2\,(\xi_2 - \xi_1)^2 + [(\overline{Z}_1 - \overline{Z}_2)\,\nu_1/w + \overline{Z}_2]\,(\xi_2 - \xi_1) - 1 = 0, \tag{15}$$

which is quadratic in $\xi_2 - \xi_1$.

On applying the quadratic formula, it follows from (15) that

$$\xi_2 - \xi_1 = \{-[(\overline{Z}_1 - \overline{Z}_2)\,\nu_1/w + \overline{Z}_2] + \sqrt{([(\overline{Z}_1 - \overline{Z}_2)\,\nu_1/w + \overline{Z}_2]^2 + 4s^2/w^2)}\}\,w^2/(2s^2), \tag{16}$$

where the positive sign is taken with the radical since $\xi_2 - \xi_1 > 0$. The first equation of (14) is solved for ξ_1 to obtain

$$-\xi_1 = [\xi_2\nu_1/w - (\overline{Z}_1 - \overline{Z}_2)]/(1 - \nu_1/w). \tag{17}$$

With this result substituted into (16), it follows that

$$\xi_2 = (\overline{Z}_1 - \overline{Z}_2) + \{[(\overline{Z}_1 - \overline{Z}_2)\,\nu_1/w + \overline{Z}_2] - \sqrt{([(\overline{Z}_1 - \overline{Z}_2)\,\nu_1/w + \overline{Z}_2]^2 + 4s^2/w^2)}\}\,w(\nu_1 - w)/(2s^2). \tag{18}$$

Since (17) and (18) are of the forms specified by (13), they may be used to iterate to the required values $\hat{\xi}_1$ and $\hat{\xi}_2$ by successive substitutions as follows:

$$\left.\begin{aligned} \xi_2^{(i+1)} &= (\overline{Z}_1^{(i)} - \overline{Z}_2^{(i)}) + \{[(\overline{Z}_1^{(i)} - \overline{Z}_2^{(i)})\,\nu_1/w + \overline{Z}_2^{(i)}] - \sqrt{([(\overline{Z}_1^{(i)} - \overline{Z}_2^{(i)})\,\nu_1/w + \overline{Z}_2^{(i)}]^2 + 4s^2/w^2)}\}\,w(\nu_1 - w)/(2s^2), \\ \xi_1^{(i+1)} &= -[\xi_2^{(i+1)}\nu_1/w - (\overline{Z}_1^{(i)} - \overline{Z}_2^{(i)})]/(1 - \nu_1/w), \end{aligned}\right\} \tag{19}$$

where $\xi_j^{(i)}$ is the ith iterant (approximation) to ξ_j.

With $\xi_1^{(0)}$ and $\xi_2^{(0)}$ determined from the chart of Fig. 1, improved approximations can be obtained by the successive application of (19). In many applications it turns out that these iterants result in a rapid advance toward the neighbourhood of the solution during the first cycle or two. Thereafter, convergence of succeeding iterants slows down as the solution is approached. This behaviour is opposite to that of Newton's method, for which convergence of successive iterants is sometimes slow for values very far removed from the solution, but is more rapid in the neighbourhood of the solution. The two methods of iteration thus complement each other, and in practical applications where a high degree of accuracy in the solution of estimating equations is required, an efficient computational procedure consists of reading initial approximations $\xi_1^{(0)}$ and $\xi_2^{(0)}$ from Fig. 1, advancing to the neighbourhood of the solution with one or more cycles of (19), and obtaining final estimates with a single cycle of Newton's method. In using Newton's method for the final cycle, estimates of the variances and covariance are obtained simultaneously without the additional computational effort which would otherwise be necessary in evaluating the second derivatives separately.

Newton's method. This method of iteration is based on Taylor series expansions of estimating equations in the vicinity of the solution. With values σ_0 and μ_0 designating approximate solutions to the estimating equations, $\partial L/\partial\mu = 0$ and $\partial L/\partial\sigma = 0$, we write

$$\mu = \mu_0 + h, \quad \sigma = \sigma_0 + k. \tag{20}$$

By Taylor's theorem, neglecting powers of h and k above the first, we have

$$\left.\begin{aligned} h\frac{\partial^2 L}{\partial \mu_0^2}+k\frac{\partial^2 L}{\partial \mu_0 \partial \sigma_0} &= -\frac{\partial L}{\partial \mu_0}, \\ h\frac{\partial^2 L}{\partial \mu_0 \partial \sigma_0}+k\frac{\partial^2 L}{\partial \sigma_0^2} &= -\frac{\partial L}{\partial \sigma_0}. \end{aligned}\right\} \tag{21}$$

Corrections h and k are then obtained as the simultaneous solution of (21).

For a doubly truncated sample, the coefficients of (21), obtained by differentiating (2), are

$$\left.\begin{aligned} \frac{\partial L}{\partial \mu} &= \frac{n}{\sigma^2}[\bar{x}-\mu-\sigma(\bar{Z}_1-\bar{Z}_2)], \\ \frac{\partial L}{\partial \sigma} &= \frac{n}{\sigma^2}[\{s^2+(\bar{x}-\mu)^2\}/\sigma-\sigma(1+\xi_1\bar{Z}_1-\xi_2\bar{Z}_2)], \\ \frac{\partial^2 L}{\partial \mu^2} &= -\frac{n}{\sigma^2}[1+\xi_1\bar{Z}_1-\xi_2\bar{Z}_2-(\bar{Z}_1-\bar{Z}_2)], \\ \frac{\partial^2 L}{\partial \mu \partial \sigma} &= -\frac{n}{\sigma^2}\left[\frac{2(\bar{x}-\mu)}{\sigma}-(\bar{Z}_1-\bar{Z}_2)(1+\xi_1\bar{Z}_1-\xi_2\bar{Z}_2)+\xi_1^2\bar{Z}_1-\xi_2^2\bar{Z}_2\right], \\ \frac{\partial^2 L}{\partial \sigma^2} &= -\frac{n}{\sigma^2}\left[\frac{3\{s^2+(\bar{x}-\mu)^2\}}{\sigma^2}-(1+\xi_1\bar{Z}_1-\xi_2\bar{Z}_2)^2+\xi_1^3\bar{Z}_1-\xi_2^3\bar{Z}_2\right], \end{aligned}\right\} \tag{22}$$

where $\bar{x}$ $(=x_0+\nu_1)$ has been written for the sample mean.

Newton's method has been set up here in terms of μ and σ rather than ξ_1 and ξ_2 as for the successive substitutions of (19). Thereby, asymptotic variances of the estimates can be readily approximated using the second partial derivatives from (22), as evaluated for the final cycle of iteration. This is permissible, since these values closely approximate corresponding expected values which would be required for more exact calculation of the variances.

4. Doubly censored samples

In this section we are concerned with doubly censored samples from a population with probability density function (1), such that out of a fixed total of N random observations, there are n_1 for which it is known only that $x<x_0$, n_2 for which it is known only that $x>x_0+w$, and $n=N-n_1-n_2$ fully measured observations in the interval $x_0\leqslant x\leqslant x_0+w$. For a sample of this type, the logarithm of the likelihood function is

$$L=n_1\ln[1-I(\xi_1)]+n_2\ln I(\xi_2)-n\ln\sigma-\sum_1^n(x_i-\mu)^2/2\sigma^2+\text{const.} \tag{23}$$

This case also was considered by Cohen (1950), and the estimating equations obtained by setting $\partial L/\partial\mu=0$ and $\partial L/\partial\sigma=0$ may be expressed as

$$\left.\begin{aligned} [Y_1-Y_2-\xi_1]/(\xi_2-\xi_1)-\nu_1/w &= 0, \\ [1+\xi_1Y_1-\xi_2Y_2-(Y_1-Y_2)^2]/(\xi_2-\xi_1)^2-s^2/w^2 &= 0, \end{aligned}\right\} \tag{24}$$

where

$$\left.\begin{aligned} Y_1(\xi_1) &= \frac{n_1}{n}\frac{\phi(\xi_1)}{1-I(\xi_1)}=\frac{n_1}{n}Z(-\xi_1), \\ Y_2(\xi_2) &= \frac{n_2}{n}\frac{\phi(\xi_2)}{I(\xi_2)}=\frac{n_2}{n}Z(\xi_2), \end{aligned}\right\} \tag{25}$$

with $Z(\xi)$ designating the reciprocal of Mill's ratio, i.e.

$$Z(\xi)=\phi(\xi)/I(\xi). \tag{26}$$

Tables of this ratio and of related functions recently appeared as an Editorial (1955) in this journal. In these tables, $Z(\xi)$ is denoted as Z/P and $Z(-\xi)$ as Z/Q, with X written for the standardized normal deviate rather than ξ.

As given above, equations (24) are analogous to those of (10) for doubly truncated samples. They differ only in the occurrence of Y_1 and Y_2 in place of $\overline{Z}_1$ and $\overline{Z}_2$ respectively.

In § 2 it was pointed out that for doubly truncated samples the method of maximum likelihood yields the same estimating equations as the method of moments. The same is true in the case of censored samples if we suppose that the n_1 and n_2 censored sample observations have the same first and second moments as the integrated moments for the tails of a normal distribution containing proportions n_1/N and n_2/N of the whole. This latter observation was made by Des Raj (1953), who described his method as one of *modified moments*.

Since Y_1 and Y_2 involve not only ξ_1 and ξ_2, but also n_1, n and n_2, charts corresponding to Fig. 1 do not seem practical in this case. We must accordingly expect to start with less accurate initial approximations than in the doubly truncated case, and depend on iteration for improving these approximations to the required accuracy.

When n_1 and n_2 constitute only a small proportion of the sample, they may be neglected to the extent that initial approximations to $\hat{\xi}_1$ and $\hat{\xi}_2$ can be read from Fig. 1 as for a doubly truncated sample. When n_1 and n_2 are appreciable proportions of the total sample, better first approximations might be read from tables of the normal curve areas using the relations

$$\frac{n_1}{n_1+n_2+n} = \int_{-\infty}^{\xi_1} \phi(t)\,dt, \quad \frac{n_2}{n_1+n_2+n} = \int_{\xi_2}^{\infty} \phi(t)\,dt. \tag{27}$$

For reasonably large samples, the two sets of initial approximations suggested above should be in fairly close agreement, and either set might be satisfactory as a starting point in the iterative process. When appreciable differences exist between the two sets of values, the choice of starting point might appropriately be based in some measure on the magnitude of n_1 and n_2 relative to the total sample size as already mentioned. Under some circumstances an average of the two might be preferred.

Iteration procedures given in § 3 for doubly truncated samples also apply here with slight modifications. The iterants of (19) are applicable in this case when Y_1 and Y_2 are substituted for $\overline{Z}_1$ and $\overline{Z}_2$ respectively. Likewise Newton's method is applicable where the coefficients of (21) are obtained from (23) by differentiation. With Y_1 and Y_2 defined by (25) and $Z(\xi)$ defined by (26), these coefficients are

$$\left.\begin{aligned}
\frac{\partial L}{\partial \mu} &= \frac{n}{\sigma^2}[\overline{x}-\mu-\sigma(Y_1-Y_2)],\\
\frac{\partial L}{\partial \sigma} &= \frac{n}{\sigma^2}\left[\frac{s^2+(\overline{x}-\mu)^2}{\sigma}-\sigma(1+\xi_1Y_1-\xi_2Y_2)\right],\\
\frac{\partial^2 L}{\partial \mu^2} &= -\frac{1}{\sigma^2}[n+n_1Z'(-\xi_1)+n_2Z'(\xi_2)],\\
\frac{\partial^2 L}{\partial \mu\,\partial \sigma} &= -\frac{1}{\sigma^2}\left[\frac{2n(\overline{x}-\mu)}{\sigma}-n_1\lambda(-\xi_1)+n_2\lambda(\xi_2)\right],\\
\frac{\partial^2 L}{\partial \sigma^2} &= -\frac{1}{\sigma^2}\left[\frac{3n\{s^2+(\overline{x}-\mu)^2\}}{\sigma^2}-n+n_1\zeta(-\xi_1)+n_2\zeta(\xi_2)\right],
\end{aligned}\right\} \tag{28}$$

where

$$\left.\begin{aligned}
Z'(\xi) &= Z(\xi)\,[Z(\xi)-\xi],\\
\lambda(\xi) &= Z(\xi)+\xi Z'(\xi),\\
\zeta(\xi) &= \xi[Z(\xi)+\lambda(\xi)].
\end{aligned}\right\} \tag{29}$$

As written above, Z' is the derivative of Z with respect to ξ.

5. Singly censored samples

For a singly censored sample in which, out of a fixed total of N observations, there are n measured observations for which $x \geqslant x_0$ and $n_1 = N - n$ unmeasured observations for which it is known only that $x < x_0$, the estimating equations as given by Cohen (1950) are

$$\left.\begin{aligned} [1-\hat{\xi}(\hat{Y}-\hat{\xi})]/(\hat{Y}-\hat{\xi})^2 - \nu_2/\nu_1^2 &= 0, \\ \hat{\sigma} &= \nu_1/(\hat{Y}-\hat{\xi}), \\ \hat{\mu} &= x_0 - \hat{\sigma}\hat{\xi}. \end{aligned}\right\} \tag{30}$$

These results follow from (24) when we let $\xi_2 \to \infty$ and subsequently drop the subscripts from ξ_1 and Y_1. The first equation of (30) can be solved for $\hat{\xi}$. Then $\hat{\sigma}$ follows from the second, with $\hat{\mu}$ being determined from the third as in the doubly restricted cases.

If, however, $(\hat{Y}-\hat{\xi})$ is eliminated between the first two equations of (30), we thereby obtain

$$\frac{1-\hat{\xi}(\nu_1/\hat{\sigma})}{(\nu_1/\hat{\sigma})^2} - \frac{\nu_2}{\nu_1^2} = 0, \tag{31}$$

and it follows that

$$\hat{\xi} = \frac{\hat{\sigma}^2 - \nu_2}{\nu_1 \hat{\sigma}}. \tag{32}$$

Upon substituting this value for $\hat{\xi}$ into the third equation of (30), we have*

$$\hat{\mu} = x_0 - (\hat{\sigma}^2 - \nu_2)/\nu_1. \tag{33}$$

Thus as an alternate procedure for calculating estimates, we might solve the first equation of (30) for $(\hat{Y}-\hat{\xi})$ rather than for $\hat{\xi}$, then calculate $\hat{\sigma}$ using the second equation of (30) as before, and calculate $\hat{\mu}$ from (33).

Since $Y(\xi) = (n_1/n)\,Z(-\xi)$ in the notation of equation (26), the *Biometrika* Editorial tables (1955) or tables of normal curve areas and ordinates permit ready evaluation of this function, and it has been possible to prepare a chart, included here as Fig. 3, of the family of curves

$$\nu_2/\nu_1^2 = [1-\xi(Y-\xi)]/(Y-\xi)^2,$$

plotted as functions of ξ for the various values of h, where

$$h = n_1/(n+n_1). \tag{34}$$

To the extent of the range covered, tables of Hald (1949) were employed in plotting these curves. For values beyond the range of Hald's tables, computations were based directly on tables of the normal curve. In using the chart of Fig. 3, ν_2/ν_1^2 and h are computed from the sample data, and with these values known, $\hat{\xi}$ is read along the horizontal scale. When more accurate results are required, $[1-\xi(Y-\xi)]/(Y-\xi)^2$ can be calculated for additional values of ξ so that a more accurate determination of $\hat{\xi}$ or of $(\hat{Y}-\hat{\xi})$ can be obtained through interpolation.

The function $(Y-\xi)$ has been tabulated by Gupta (1952) as a function of p and ψ, where

$$p = 1-h \quad \text{and} \quad \psi = s^2/\nu_2.$$

Corresponding to observed values of p and ψ, $(Y-\xi)$, which Gupta denotes as z, is obtained by interpolation from his table. Estimates of the population standard deviation and mean

* This result was given in an equivalent form by Gupta (1952). Expressed in the notation of this paper, his estimator may be written as $\hat{\mu} = \bar{x} + (\hat{\sigma}^2 - s^2)/(\bar{x} - x_0)$.

are then calculated from the second equation of (30) and from (33) respectively. Gupta's procedure accordingly enjoys the advantage of requiring interpolation in only one table, but at any rate in parts of his table his tabular intervals seem too large for easy interpolation.*

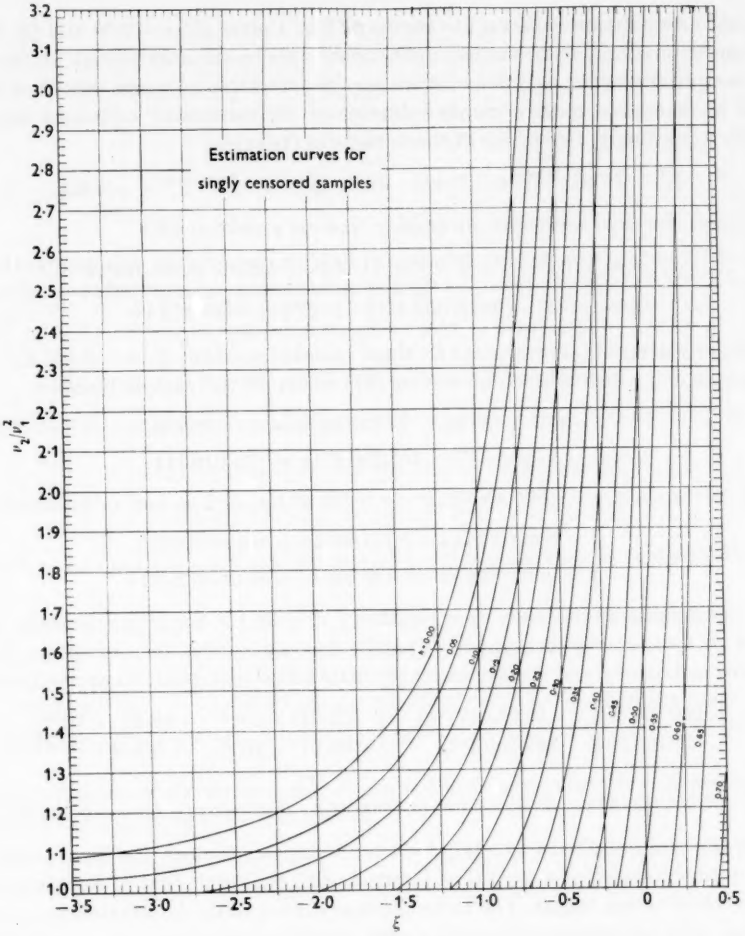


Fig. 3

6. ILLUSTRATIVE EXAMPLES

Doubly truncated sample. To illustrate estimation in the doubly truncated case, we consider an example in which the entire production of a certain bushing is sorted through go, no-go gauges, with the result that items of diameter in excess of 0·6015 in. and those

[* It seems to be worth considering whether this disadvantage might be overcome by an expansion of Gupta's table, using finer tabular intervals. Ed.]

less than 0·5985 in. are discarded. For a random sample of 75 bushings selected from the screened production,

$$\bar{x} = 0{\cdot}600\,149\,33 \text{ in.}, \quad s^2 = 0{\cdot}000\,000\,371\,187, \quad x_0 = 0{\cdot}5985 \quad \text{and} \quad w = 0{\cdot}0030.$$

Thus $$\nu_1 = \bar{x} - x_0 = 0{\cdot}001\,649\,33, \quad \nu_1/w = 0{\cdot}549\,78, \quad s^2/w^2 = 0{\cdot}041\,242,$$

and visual interpolation between the curves of Fig. 1 gives $\xi_1^{(0)} = -2{\cdot}50$ and $\xi_2^{(0)} = 2{\cdot}00$ as initial approximations. These values might conceivably be accurate enough for the purpose of this sample. However, as a demonstration of the iterative processes described in § 3, we proceed to determine more accurate solutions of the estimating equations using those methods. Accordingly, two cycles of the iterants of (19) yield

$$\xi_2^{(1)} = 1{\cdot}997, \quad \xi_1^{(1)} = -2{\cdot}522 \quad \text{and} \quad \xi_2^{(2)} = 1{\cdot}997, \quad \xi_1^{(2)} = -2{\cdot}525,$$

respectively. For final estimates, we employ Newton's method with

$$\sigma_0 = w/(\xi_2^{(2)} - \xi_1^{(2)}) = 0{\cdot}0030/(1{\cdot}997 + 2{\cdot}525) = 0{\cdot}000\,663\,42$$

and $$\mu_0 = 0{\cdot}5985 - (0{\cdot}000\,663\,42)(-2{\cdot}525) = 0{\cdot}600\,175\,14.$$

With the derivatives of (22) evaluated for these values of μ_0 and σ_0 ($\xi_1 = -2{\cdot}525, \xi_2 = 1{\cdot}997$), corrections h and k are obtained by solving (21) which for this sample becomes

$$\begin{aligned} -150\,706\,576h + \ 27\,235\,297k &= -7{\cdot}076\,300, \\ 27\,235\,297h - 187\,700\,371k &= \ 75{\cdot}019\,711. \end{aligned}$$

Thereby we obtain $h = -0{\cdot}000\,000\,03$, $k = -0{\cdot}000\,000\,40$, and as final estimates we have

$$\begin{aligned} \hat{\mu} &= 0{\cdot}600\,175\,14 - 0{\cdot}000\,000\,03 = 0{\cdot}600\,175\,11, \\ \hat{\sigma} &= 0{\cdot}000\,663\,42 - 0{\cdot}000\,000\,40 = 0{\cdot}000\,663\,02. \end{aligned}$$

Since coefficients of the correction equations in h and k are approximately equal to expected values of corresponding second partial derivatives of L, the asymptotic variance-covariance matrix of $\hat{\mu}$ and $\hat{\sigma}$ may, with very little additional effort, be approximated as

$$\begin{pmatrix} 150\,706\,576 & -27\,235\,297 \\ -27\,235\,297 & 187\,700\,371 \end{pmatrix}^{-1} = \begin{pmatrix} 5{\cdot}471 \times 10^{-9} & -98{\cdot}87 \times 10^{-9} \\ -98{\cdot}87 \times 10^{-9} & 6{\cdot}814 \times 10^{-9} \end{pmatrix}.$$

Thus,

$$V(\hat{\mu}) \sim 0{\cdot}000\,000\,005\,471, \quad V(\hat{\sigma}) \sim 0{\cdot}000\,000\,006\,814 \quad \text{and} \quad \operatorname{cov}(\hat{\mu}, \hat{\sigma}) \sim -0{\cdot}000\,000\,098\,87.$$

Doubly censored sample. In a certain time-mortality experiment, the first observation is delayed until the elapse of a fixed time interval, with the result that n_1 sample specimens die before observation begins. The experiment is subsequently terminated at a predetermined time with n_2 specimens remaining alive. Actual survival times are recorded for the n specimens which die during the period of full observation. For a specific sample of this type in which x designates the logarithm of survival time in days and is assumed to be normally distributed (μ, σ),

$$n_1 = 2, \quad n = 40, \quad n_2 = 5, \quad x_0 = 1{\cdot}301\,030, \quad w = 0{\cdot}602\,060, \quad \bar{x} = 1{\cdot}620\,111,$$

$$\nu_1 = 0{\cdot}319\,081, \quad s^2 = 0{\cdot}021\,739\,2, \quad \nu_1/w = 0{\cdot}529\,982 \quad \text{and} \quad s^2/w^2 = 0{\cdot}059\,974\,1.$$

Neglecting knowledge of n_1 and n_2, we read $\xi_1 = -1{\cdot}71$ and $\xi_2 = 1{\cdot}36$ from Fig. 1. Subsequently we apply (27) to obtain $\xi_1 = -1{\cdot}72$ and $\xi_2 = 1{\cdot}25$ from tables of normal curve

areas. The two sets of readings are in reasonably close agreement, and we decide to begin the iteration with initial values of $\xi_1^{(0)} = -1{\cdot}72$, $\xi_2^{(0)} = 1{\cdot}28$. The iterants of (19) with $\bar{Z}_1$ and $\bar{Z}_2$ replaced by Y_1 and Y_2 respectively are employed to obtain the results tabulated below.

i	$\xi_1^{(i)}$	$\xi_2^{(i)}$
0	−1·72	1·28
1	−1·681	1·278
2	−1·689	1·282
3	−1·688	1·281
4	−1·688	1·281

As a demonstration of Newton's method in this case, we start with the results obtained above at the end of the second cycle, that is, with $\xi_1^{(2)} = -1{\cdot}689$ and $\xi_2^{(2)} = 1{\cdot}282$. Accordingly

$$\sigma_0 = 0{\cdot}602\,060/(1{\cdot}282 + 1{\cdot}689) = 0{\cdot}202\,646$$

and

$$\mu_0 = 1{\cdot}301\,030 - (0{\cdot}202\,646)(-1{\cdot}689) = 1{\cdot}643\,299.$$

With derivatives of (28) calculated for these values of μ_0 and σ_0, equations (21) become

$$-1117{\cdot}362\,555h + \quad 52{\cdot}976\,746k = \quad 0{\cdot}009\,567\,35,$$

$$52{\cdot}976\,746h + 1791{\cdot}254\,507k = -0{\cdot}234\,567\,59.$$

On solution, we obtain $h = -0{\cdot}000\,002\,36$ and $k = 0{\cdot}000\,130\,89$, and as final estimates we have

$$\hat{\mu} = 1{\cdot}643\,299 - 0{\cdot}000\,002 = 1{\cdot}643\,297,$$

$$\hat{\sigma} = 0{\cdot}202\,646 + 0{\cdot}000\,131 = 0{\cdot}202\,777,$$

which correspond to $\hat{\xi}_1 = -1{\cdot}68790$ and $\hat{\xi}_2 = 1{\cdot}28122$. It is to be noted that when rounded off to three decimals, these two latter values agree with those obtained using the iterants of (19).

Using coefficients of the above correction equations, the asymptotic variance-covariance matrix of $\hat{\mu}$ and $\hat{\sigma}$ may be approximated as

$$\begin{pmatrix} 1117{\cdot}3626 & -52{\cdot}9767 \\ -52{\cdot}9767 & 1791{\cdot}2545 \end{pmatrix}^{-1} = \begin{pmatrix} 0{\cdot}000\,559\,1 & -0{\cdot}000\,026\,5 \\ -0{\cdot}000\,026\,5 & 0{\cdot}000\,896\,2 \end{pmatrix}.$$

Thus we have,

$$V(\hat{\mu}) \sim 0{\cdot}000\,559\,1, \quad V(\hat{\sigma}) \sim 0{\cdot}000\,896\,2 \quad \text{and} \quad \operatorname{cov}(\hat{\mu}, \hat{\sigma}) \sim -0{\cdot}000\,026\,5.$$

Singly censored sample. To illustrate estimation from a singly censored sample, we consider one for which $x_0 = 70{\cdot}00$, $n = 50$, $n_1 = 3$, $\nu_1 = 10{\cdot}654$, $\nu_2 = 145{\cdot}2426$. Thus, $h = 3/(3+50) = 0{\cdot}05664$ and $\nu_2/\nu_1^2 = 1{\cdot}27958$. Reading from Fig. 3 we have $\xi = -1{\cdot}56$ as a first approximation. Since the computing routine is simple, the accuracy of this result can readily be improved by calculating additional values of $[1 - \xi(Y - \xi)]/(Y - \xi)^2$ using a table of normal curve areas and ordinates, and then interpolating as summarized below.

When the *Biometrika* Editorial tables (1955) are available, the required values of $Z(-\xi)$ can be read directly without the necessity of the additional calculations involved in using normal curve tables.

ξ	$Y-\xi$	$[1-\xi(Y-\xi)]/(Y-\xi)^2$
$-1{\cdot}560$	1·67939	1·28347
$-1{\cdot}569$	*1·68899*	1·27958
$-1{\cdot}570$	1·68990	1·27921

With $\hat{\xi} = -1{\cdot}569$ and $(\hat{Y}-\hat{\xi}) = 1{\cdot}68899$, we estimate σ and μ from the second and third equations of (30) as

$$\hat{\sigma} = 10{\cdot}654/1{\cdot}68899 = 6{\cdot}308,$$

$$\hat{\mu} = 70{\cdot}00-(-1{\cdot}569)(6{\cdot}308) = 79{\cdot}90.$$

Alternately, we might have estimated μ from (33) to obtain the same value as above

$$\hat{\mu} = 70{\cdot}00-(6{\cdot}308^2-145{\cdot}2426)/10{\cdot}654 = 79{\cdot}90.$$

REFERENCES

Cohen, A. C., Jr. (1950). Estimating the mean and variance of normal populations from singly truncated and doubly truncated samples. *Ann. Math. Statist.* **21**, 557.

Cohen, A. C., Jr. & Woodward, John (1953). Tables of Pearson–Lee–Fisher functions of singly truncated normal distributions. *Biometrics*, **9**, 489.

Editorial (1955). The normal probability function: Tables of certain area-ordinate ratios and their reciprocals. *Biometrika*, **42**, 217.

Gupta, A. K. (1952). Estimation of the mean and standard deviation of a normal population from a censored sample. *Biometrika*, **39**, 260.

Hald, A. (1949). Maximum likelihood estimation of the parameters of a normal distribution which is truncated at a known point. *Skand. AktuarTidskr.* **32**, 119.

Norton, H. W. (1956). One likelihood adjustment may be inadequate. *Biometrics*, **12**, 79.

Raj, Des (1953). On moments estimation of the parameters of a normal population from singly and doubly truncated samples. *Ganita*, **4**, 79.

Scarborough, J. B. (1930). *Numerical Mathematical Analysis*. Baltimore: Johns Hopkins Press.

Thomson, G. W., Friedman, M. & Garelis, E. (1954). Estimation of mean and variance of normal populations from doubly truncated samples. *Report of Ethyl Corporation Research Laboratories*, Detroit.

UPPER PERCENTAGE POINTS OF THE GENERALIZED BETA DISTRIBUTION. I

By F. G. FOSTER and D. H. REES

Research Techniques Unit, London School of Economics and Rothamsted Experimental Station, Harpenden

1. Introduction

In sampling from k-variate Normal populations, if A and B are independent estimates, based on ν_1 and ν_2 degrees of freedom, of what on a null hypothesis is the same dispersion matrix, the roots of the determinantal equation

$$|B-\lambda A| = 0$$

are invariant under all linear transformations of the k variates. They are thus unaffected, for example, by a change in the unit of measurement, and so are independent of the magnitudes of the parent variances and covariances. There are clearly advantages in testing the null hypothesis by means of some function of the roots, and arguments for their use have been advanced (Hotelling, 1947, 1951, 1954; Roy, 1945, 1950, 1953, 1954). Among the functions proposed* are: (i) the product of the roots (Wilks, 1932; Pearson & Wilks, 1933); (ii) the sum of the roots (Hotelling, 1947, 1951); (iii) the greatest (or least) root (Roy, 1950, 1953, 1954). We are concerned here with the tabulation of the greatest root, for the case, $k = 2$. The computations are based on a recursion formula of Roy (1945). Similar tabulations for the cases $k = 3$, 4 and 5 are in hand.

If A is based on ν_1 and B on ν_2 degrees of freedom, the k roots of

$$|\nu_2 B-\theta(\nu_1 A+\nu_2 B)| = 0$$

all lie between 0 and 1 and are related to the roots λ of the previous equation by

$$\theta = \frac{\nu_2\lambda}{\nu_1+\nu_2\lambda}.$$

Thus the roots λ are immediately obtainable from the roots θ and vice versa. The relation corresponds to that which exists between the F and the Beta distributions in the univariate case. We have tabulated the 80, 85, 90, 95 and 99 % points of the distribution of the greatest root θ on the null hypothesis of identical parent dispersion matrices. The tables extend previous tables by Pillai (1956) and Nanda (1951).

The joint distribution of the k roots, $\theta_1, \theta_2, \ldots, \theta_k$, has been given by Fisher (1939), Hsu (1939), Roy (1939) and Girshick (1939) (cf. also Mood, 1951). The density function is

$$K\prod_{i=1}^{k}\theta_i^{p-1}(1-\theta_i)^{q-1}\prod_{i>j}(\theta_i-\theta_j) \quad (0<\theta_1\leqslant\ldots\leqslant\theta_k<1),$$

where

$$K = \pi^{\frac{1}{2}k}\prod_{i=1}^{k}\frac{\Gamma(\frac{1}{2}\overline{2p+2q+k+i-2})}{\Gamma(\frac{1}{2}2p+i-1)\,\Gamma(\frac{1}{2}2q+i-1)\,\Gamma(\frac{1}{2}i)},$$

and

$$p = \tfrac{1}{2}(\nu_2-k+1), \quad q = \tfrac{1}{2}(\nu_1-k+1).$$

* The question as to the relative merits of these functions in different circumstances does not appear to have been fully resolved, and will not be taken up here. A possible approach will be to carry out distribution sampling experiments to obtain the power functions of the tests against various alternatives.

Thus the distribution function of the greatest root is given by

$$I_x(k;\, p, q) = K\int_0^x d\theta_k \int_0^{\theta_k} d\theta_{k-1} \dots \int_0^{\theta_2} d\theta_1 \prod_{i=1}^{k} \theta_i^{p-1}(1-\theta_i)^{q-1} \prod_{i>j} (\theta_i - \theta_j).$$

It will be seen that $I_x(k;\, p, q)$ reduces to the Beta distribution $I_x(p, q)$ when $k = 1$, and for this reason we have proposed the name 'Generalized Beta distribution' for it.

2. Interpolation

For use of the tables in connexion with tests of significance, the value of x will be required for the percentage levels tabulated and for integer values of $\nu_2 = 2p+1$, $\nu_1 = 2q+1$. Linear interpolation, either p-wise or q-wise, within the body of the tables will usually suffice for accuracy to at least two decimal places. For greater accuracy 3- or 4-point Lagrangian interpolation should be used. Extrapolation beyond $\nu_2 = 21$ ($p = 10$) is not possible, and the necessity for this is not expected to arise. For q-wise interpolation between $\nu_1 = 161$ ($q = 80$) and $\nu_1 = \infty$, 3-point harmonic interpolation, based on $q = 40$, $q = 80$, $q = \infty$ may be used.

Example. Find the 95 % point corresponding to $\nu_2 = 13$ ($p = 6$) and $\nu_1 = 403$ ($q = 201$):

ν_1	q	$1/q$	x	Δx	$\Delta^2 x$
∞	∞	0	0		
				1506	
161	80	1/80	0·1506		−304
				1202	
81	40	1/40	0·2708		

The values, $1/q$, are equally spaced, with interval 1/80, and we require the value of x corresponding to $1/q = 1/201$. Thus

$$\begin{aligned} x &= 0 + \frac{1/201}{1/80}\Delta x_0 + \frac{(1/201)(1/201 - 1/80)}{2(1/80)^2}\Delta^2 x_0 \\ &= \frac{80}{201}(0{\cdot}1506) + \frac{80 \times 121}{2 \times 201^2}(0{\cdot}0304) \\ &= 0{\cdot}064, \end{aligned}$$

which is correct to 3 decimal places.

3. Uses of the tables

We illustrate two typical applications to tests of significance discussed by Roy (1945, 1950): the analysis of dispersion (i) of means and (ii) of regression. In both cases the appropriate critical region is the upper tail of the distribution (cf. Rao, 1952, chapter 7).

It should be noted that a direct application to testing the *equality* of two dispersion matrices, when we do not know which should be the 'larger', would require the joint distribution of the greatest and least roots. This would correspond to the univariate test in which we use both tails of the F (or Beta) distribution as critical region. Since the greatest and least roots are not independently distributed, these tables are not appropriate for this type of test.

(i) *Analysis of dispersion of means*

Let us select samples of sizes $n_1, n_2, \ldots, n_r$ from r bivariate Normal populations, having the same dispersion matrix. It is required to test the hypothesis that the r mean vectors of the populations are equal. Let $n_1+n_2+\ldots+n_r = N$. The combined sums of products matrix (S) based on $N-1$ degrees of freedom, may be analysed into sums of products due to the means (Q) with $r-1$ degrees of freedom and sums of products due to the error (W) with $N-r$ degrees of freedom:

$$S = Q+W.$$

Then Q and W, divided by their degrees of freedom, are independent estimates of the same parent dispersion matrix, and our test consists in computing the greatest root of $|Q-\theta S| = 0$, and entering the table with $\nu_1 = N-r$, $\nu_2 = r-1$.

Example. Table 1 gives the analysis of dispersion for the two characters $x_1 = \log_{10}$ (tooth length) and $x_2 = \log_{10}$ (maximum breadth) of the permanent upper second premolar for the three male groups, human (West African), chimpanzee and orang-outang. The groups have been selected from a larger body of data kindly made available to us by the authors (Ashton, Healy & Lipton, 1957). The logarithmic transformation was used to stabilize the within-group variance. Then

$$Q = \begin{pmatrix} 0{\cdot}544941 & 0{\cdot}525768 \\ 0{\cdot}525768 & 0{\cdot}509075 \end{pmatrix} \quad \text{and} \quad S = \begin{pmatrix} 0{\cdot}682727 & 0{\cdot}595110 \\ 0{\cdot}595110 & 0{\cdot}601867 \end{pmatrix}.$$

Table 1

	Degrees of freedom	Sums of products matrix			
			x_1^2	x_1x_2	x_2^2
Between groups	2	Q_{ij}	0·544941	0·525768	0·509075
Within groups	154	W_{ij}	0·137786	0·069342	0·092792
Total	156	S_{ij}	0·682727	0·595110	0·601867

The roots of $|Q-\theta S| = 0$ are now obtained by solving the quadratic

$$(0{\cdot}544941-\theta\,0{\cdot}682727)(0{\cdot}509075-\theta\,0{\cdot}601867) - (0{\cdot}525768-\theta\,0{\cdot}595110)(0{\cdot}525768-\theta\,0{\cdot}595110) = 0.$$

We find that $\theta_1 = 0{\cdot}020238, \quad \theta_2 = 0{\cdot}856543.$

With $\nu_2 = 2$ and $\nu_1 = 154$, $\theta_{\max.} = 0{\cdot}857$ is seen to be highly significant, since from the tables we see that the 99 % point for $\nu_2 = 2$, $\nu_1 = 121$ is 0·096, and for $\nu_2 = 2$, $\nu_1 = 161$ it is 0·073.

The group means are given in Table 2.

For the sake of comparison we illustrate the alternative test criterion based on the product of the roots, using these data. The statistic $\Lambda = |W|/|S|$ is equal to the product $\prod_{i=1}^{k}(1-\theta_i)$. This criterion is arrived at by the likelihood ratio method by Pearson & Wilks (1933), and for $k = 2$, $\sqrt{\Lambda}$ has exactly the Beta distribution $I_x(\nu_1 - 1, \nu_2)$ on the null hypothesis. Applied to the above data, this gives $\Lambda = 0{\cdot}140554$. Therefore $\sqrt{\Lambda} = 0{\cdot}375$, and on the null hypothesis that neither x_1 nor x_2 differs from group to group, $\sqrt{\Lambda}$ should be distributed according to $I_x(153, 2)$. The lower tail of the distribution being here the critical region, this is again highly significant, since $\Pr(\sqrt{\Lambda} < 0{\cdot}375) = 0{\cdot}65 \times 10^{-63}$. (Another example is given by Pearson & Wilks, 1933, p. 370.)

Table 2

	No. in group	$\bar{x}_1$	$\bar{x}_2$
Human (West African)	59	1·846	1·981
Chimpanzee	55	1·865	2·008
Orang-outang	43	1·986	2·119

(ii) *Analysis of dispersion of regression*

Let $x_1, x_2, \ldots, x_s, x_{s+1}, x_{s+2}$ be $(s+2)$ correlated variables for which a combined sums-of-products matrix (x_{ij}), based on ν degrees of freedom, is obtained. It is required to test the hypothesis that the linear regression of x_{s+1}, x_{s+2} on the other s variables,

$$\mathscr{E}(x_{s+1}) = \sum_{i=1}^{s} b_{i1} x_i,$$

$$\mathscr{E}(x_{s+2}) = \sum_{i=1}^{s} b_{i2} x_i,$$

has zero regression coefficients, i.e. that the variation in x_{s+1}, x_{s+2} is independent of that due to the other variables.

The sums of products matrix for $x_1, \ldots, x_s$ is

$$\begin{pmatrix} x_{11} & \ldots & x_{1s} \\ \vdots & & \\ x_{s1} & \ldots & x_{ss} \end{pmatrix}.$$

Let the inverse be

$$\begin{pmatrix} x^{11} & \ldots & x^{1s} \\ \vdots & & \\ x^{s1} & \ldots & x^{ss} \end{pmatrix}.$$

Then the sums-of-products matrix for x_{s+1}, x_{s+2} due to $x_1, \ldots, x_s$ is

$$Q = \begin{pmatrix} x_{1,s+1} & \ldots & x_{s,s+1} \\ x_{1,s+2} & \ldots & x_{s,s+2} \end{pmatrix} \begin{pmatrix} x^{11} & \ldots & x^{1s} \\ \vdots & & \\ x^{s1} & \ldots & x^{ss} \end{pmatrix} \begin{pmatrix} x_{1,s+1} & x_{1,s+2} \\ \vdots & \\ x_{s,s+1} & x_{s,s+2} \end{pmatrix}.$$

On the null hypothesis Q yields an estimate, with s degrees of freedom, of the dispersion matrix of x_{s+1}, x_{s+2}. The sum of products matrix for x_{s+1}, x_{s+2} due to the error, after correcting for $x_1, \ldots, x_s$, is now

$$W = \begin{pmatrix} x_{s+1,s+1} & x_{s+1,s+2} \\ x_{s+2,s+1} & x_{s+2,s+2} \end{pmatrix} - Q.$$

This yields an estimate based on $\nu - s$ degrees of freedom, which, on the null hypothesis, is independent of the previous one. Therefore

$$S = \begin{pmatrix} x_{s+1,s+1} & x_{s+1,s+2} \\ x_{s+2,s+1} & x_{s+2,s+2} \end{pmatrix},$$

and our test consists in computing $\theta_{\max.}$ for

$$|Q - \theta S| = 0,$$

and entering the table with $\nu_1 = \nu - s$, $\nu_2 = s$.

Example. The following five measurements were made on the permanent upper first incisor of 83 female gorillas.

y_1 = basal width,
y_2 = maximum width,
y_3 = basal thickness,
y_4 = maximum height,
y_5 = thickness at maximum width.

This data is again taken from the same source acknowledged in the previous example.

For the reason given there the logarithmic transforms of these measurements, $x_i = \log_{10} y_i$, were taken and the following sums-of-products matrix (divided by 82) obtained:

$$10^{-4}\begin{array}{c} \\ x_1 \\ x_2 \\ x_3 \\ x_4 \\ x_5 \end{array}\begin{array}{c} \begin{array}{ccccc} x_1 & x_2 & x_3 & x_4 & x_5 \end{array} \\ \begin{bmatrix} 13{\cdot}03 & & & & \\ 5{\cdot}77 & 12{\cdot}36 & & & \\ 4{\cdot}90 & 8{\cdot}33 & 11{\cdot}88 & & \\ 3{\cdot}83 & 39{\cdot}14 & 28{\cdot}38 & 229{\cdot}36 & \\ -1{\cdot}95 & -44{\cdot}75 & -30{\cdot}95 & -261{\cdot}52 & 388{\cdot}31 \end{bmatrix} \end{array}.$$

It is required to test the significance of the linear regressions:

$$\mathscr{E}(x_4) = b_{14}x_1 + b_{24}x_2 + b_{34}x_3,$$
$$\mathscr{E}(x_5) = b_{15}x_1 + b_{25}x_2 + b_{35}x_3.$$

The inverse of the sums-of-products matrix for x_1, x_2 and x_3 is

$$10^4\begin{bmatrix} 0{\cdot}098298 & & \\ -0{\cdot}035196 & 0{\cdot}165996 & \\ -0{\cdot}015965 & -0{\cdot}101876 & +0{\cdot}162152 \end{bmatrix} = A^{-1}.$$

The sums-of-products matrix due to the regressions is given by

$$10^{-8}\begin{bmatrix} 3{\cdot}83 & 39{\cdot}14 & 28{\cdot}38 \\ -1{\cdot}95 & -44{\cdot}75 & -30{\cdot}95 \end{bmatrix} A^{-1} \begin{bmatrix} 3{\cdot}83 & -1{\cdot}95 \\ 39{\cdot}14 & -4{\cdot}75 \\ 28{\cdot}38 & -30{\cdot}95 \end{bmatrix} = 10^{-4}\begin{bmatrix} 146{\cdot}01 & \\ -169{\cdot}64 & 197{\cdot}86 \end{bmatrix} = Q.$$

and has 3 D.F.

The total sums-of-products matrix before the regression is

$$S = 10^{-4}\begin{bmatrix} 229{\cdot}36 & \\ -261{\cdot}52 & 388{\cdot}31 \end{bmatrix} \quad \text{with} \quad 82 \text{ D.F.}$$

The greatest latent root of $|Q-\theta S| = 0$ is $\theta_{\max.} = 0{\cdot}637$. With $\nu_2 = 3$ and $\nu_1 = 79$ this is seen to be highly significant, since from the tables we see that the 99% point for $\nu_2 = 3$ and $\nu_1 = 71$ is 0·186, and for $\nu_2 = 3$, $\nu_1 = 81$ it is 0·165.

4. Method of computation

The computations were carried out on the N.R.D.C. Elliott 401 Computer at Rothamsted Experimental Station. We restrict consideration below to the case $k = 2$ for which the tabulations have been made. Let $\theta_{\max.}$ denote the greatest root of

$$|\nu_2 B - \theta(\nu_1 A + \nu_2 B)| = 0,$$

where A and B are independent estimates, based on ν_1 and ν_2 degrees of freedom, of a parent dispersion matrix of a bivariate Normal population. Define

$$I_x(2; p, q) = \Pr\{\theta_{\max.} \leqslant x\},$$

where $p = \frac{1}{2}(\nu_2 - 1)$, $q = \frac{1}{2}(\nu_1 - 1)$. Pillai (1956) gives a formula which may be written

$$I_x(2; p, q) = K\{2B_x(2p, 2q) - x^p(1-x)^q B_x(p, q)\},$$

where

$$K = \frac{\pi^{\frac{1}{2}}\Gamma(p+q)\,\Gamma(p+q+\frac{1}{2})}{\Gamma(p)\,\Gamma(p+\frac{1}{2})\,\Gamma(q)\,\Gamma(q+\frac{1}{2})},$$

and $B_x(p, q)$ is the Incomplete Beta function. As it stands, this formula is not suitable for computation. If, however, we distribute the normalizing constant, K, we obtain the formula

$$I_x(2; p, q) = I_x(2p, 2q) - I_x(p, q)\,a_x(p, q),$$

where $I_x(p, q)$ is the Beta distribution, and

$$a_x(p, q) = x^p(1-x)^q \frac{\pi^{\frac{1}{2}}\Gamma(p+q+\frac{1}{2})}{\Gamma(p+\frac{1}{2})\,\Gamma(q+\frac{1}{2})}.$$

$I_x(p, q)$ may be computed recursively for integral values of p and q by means of the relations

$$\begin{aligned} &I_x(0, q) = 1 && (q = 0, 1, 2, \ldots), \\ &I_x(p, 0) = 0 && (p > 0), \\ &I_x(p, q) = xI_x(p-1, q) + (1-x)\,I(p, q-1) && (p, q > 0), \end{aligned}$$

and $a_x(p, q)$ may be computed by means of

$$\begin{aligned} &a_x(0, 0) = 1, \\ &a_x(p, 0) = xa_x(p-1, 0) && (p > 0), \\ &a_x(0, q) = (1-x)\,a_x(0, q-1) && (q > 0), \\ &a_x(p, q) = \left\{1 + \frac{1}{2(p+q-1)}\right\}\{xa_x(p-1, q) + (1-x)\,a_x(p, q-1)\} && (p, q > 0). \end{aligned}$$

By means of these relations, we computed tables of the distribution function $I_x(2; p, q)$ for $x = 0{\cdot}01\,(0{\cdot}01)\,0{\cdot}99$ and for the ranges $p = 1\,(1)\,10$, $q = 2\,(1)\,20\,(5)\,50$, 60, 80. A copy of these tables is available in the Library of the Royal Statistical Society. The percentage points given below were obtained, in the main, by inverse interpolation on these tables, by means of a desk calculating machine. The values have an error of less than two units in the fourth decimal place, and in general* this error is less than 0·5. Separate computations were made later to obtain the values corresponding to $p = 0{\cdot}5$.

Since for fixed x and p, $I_x(p, q) \to 1$ as $q \to \infty$, we have the approximation for small p,

$$I_x(2; p, q) \doteqdot 1 - a_x(p, q).$$

Within the range of p and q used, we found empirically that this formula was applicable when $I_x(2; p, q)$ was greater than about 0·92. The computation was switched to the approximation whenever it became applicable, which substantially increased the speed.

It will be seen from the above relations that the natural way to compute $I_x(2; p, q)$ would be not for the range of x for each p and q, as is required for tables of the distribution function, but for fixed x recursively through for p and q. To have obtained a final tabulation of the distribution function by this method we should have had to store the information consisting of about 26,000 numbers. This was beyond the internal capacity of the machine, and examination of a scheme to output the information for subsequent editing on the machine showed that the method would have seriously reduced the advantage of the more direct method of computation.

The tables were therefore prepared on the basis of a fixed q for all p and then all x. The disadvantage that the same values of $I_x(p, q)$ and $a_x(p, q)$ have to be repeatedly computed was to some extent offset by the gain in flexibility of the programme in that (i) the calculations could be restarted at any values of x, p, q, in case of any machine failure and (ii) it was possible to modify the programme easily at a later stage in order to check directly the accuracy of the percentage points obtained by inverse interpolation in the tables.

The tables were printed to four decimal places directly on the typewriter output in a normal table format: each page consisting of the distribution for one value of q, arranged in ten columns corresponding to $p = 1\,(1)\,10$ and with $x = 0{\cdot}01\,(0{\cdot}01)\,0{\cdot}99$ in groups of five. To economize on the output time, the following controls were included in the programme. The fact that $I_x(2; p, q)$ is monotonically decreasing for increasing p was used to avoid printing zeros as soon as $I_x(2; p, q)$ became less than 0·00005, nothing more was then printed on that line and a fresh calculation started with x increased by 0·01. Whenever $I_x(2; 10, q)$ became greater than 0·99995 the calculation was re-started with the next value of q and $x = 0{\cdot}01$.

5. FURTHER WORK

In the light of the experience gained with this computation, we have decided to forego the tabulations of the distribution functions in the projected computations for $k = 3$, 4 and 5, but to obtain the percentage points directly, in one automatic computation, utilizing the natural order of computation and using a method of inverse interpolation.

These tables were made in consultation with Prof. S. N. Roy, for whose advice the authors wish to express their thanks.

* The largest error is at the 99% points for $q < 20$.

REFERENCES

ASHTON, E. H., LIPTON, S. & HEALY, M. J. R. (1957). The descriptive use of discriminant functions in physical anthropology. *Proc. Roy. Soc.* B (in the Press).

FISHER, R. A. (1939). The sampling distribution of some statistics obtained from non-linear equations. *Ann. Eugen., Lond.*, **9**, 238–49. Also published in *Contributions to Mathematical Statistics*, paper 36, 1950. New York: John Wiley and Sons.

GIRSHICK, M. A. (1939). On the sampling theory of the roots of determinantal equations. *Ann. Math. Statist.* **10**, 203–24.

HOTELLING, H. (1947). Multivariate quality control illustrated by air testing of sample bombsights. From *Selected Techniques of Statistical Analysis*, ed. Eisenhart *et al.* New York.

HOTELLING, H. (1951). A generalized T test and measure of multivariate dispersion. *Proceedings of the Second Berkeley Symposium on Mathematical Statistics and Probability*. Berkeley: University of California Press.

HOTELLING, H. (1954). Multivariate analysis. From *Statistics and Mathematics in Biology*, ed. Kempthorne *et al.* Iowa: Iowa State College Press.

HSU, P. L. (1939). On the distribution of roots of certain determinantal equations. *Ann. Eugen., Lond.*, **9**, 250–8.

MOOD, A. M. (1951). On the distribution of the characteristic roots of normal second-moment matrices. *Ann. Math. Statist.* **22**, 266–73.

NANDA, D. N. (1951). Probability distribution tables of the largest root of a determinantal equation with two roots. *J. Indian Soc. Agric. Statist.* **3**, 175–7.

PEARSON, E. S. & WILKS, S. S. (1933). Methods of statistical analysis appropriate for k samples of two variables. *Biometrika*, **25**, 353–78.

PILLAI, K. C. S. (1956). On the distribution of the largest or the smallest root of a matrix in multivariate analysis. *Biometrika*, **43**, 122–7.

RAO, C. R. (1952). *Advanced Statistical Method in Biometric Research*. New York: John Wiley and Sons.

ROY, S. N. (1939). p-statistics, or some generalization in analysis of variance appropriate to multivariate problems. *Sankhyā*, **4**, 381–96.

ROY, S. N. (1945). The individual sampling distribution of the maximum, the minimum, and any intermediate of the p-statistics on the null hypothesis. *Sankhyā*, **7**, 133–58.

ROY, S. N. (1950). On some aspects of statistical inference. *Proc. International Congress of Mathematicians*, pp. 555–64.

ROY, S. N. (1953). On a heuristic method of test construction and its use in multivariate analysis. *Ann. Math. Statist.* **24**, 220–38.

ROY, S. N. (1954). A report on some aspects of multivariate analysis. Mimeo. Series no. **121**, Institute of Statistics, University of North Carolina.

WILKS, S. S. (1932). Certain generalizations in the analysis of variance. *Biometrika*, **24**, 471–94.

Generalized Beta distribution: $100P\%$ points for x

ν_1	P	$\nu_2=2$	3	5	7	9	11	13	15	17	19	21
5	0·80	0·7011	0·7728	0·8454	0·8825	0·9052	0·9205	0·9315	0·9398	0·9464	0·9516	0·9559
	·85	·7449	·8075	·8696	·9013	·9205	·9334	·9427	·9497	·9552	·9596	·9632
	·90	·7950	·8463	·8968	·9221	·9374	·9476	·9550	·9605	·9649	·9683	·9712
	·95	·8577	·8943	·9296	·9471	·9576	·9645	·9696	·9733	·9763	·9787	·9806
	·99	·9377	·9542	·9698	·9774	·9819	·9850	·9872	·9888	·9900	·9910	·9918
7	0·80	0·5638	0·6469	0·7416	0·7954	0·8303	0·8550	0·8734	0·8876	0·8989	0·9082	0·9158
	·85	·6085	·6851	·7712	·8194	·8507	·8726	·8889	·9014	·9114	·9196	·9264
	·90	·6628	·7307	·8058	·8474	·8741	·8928	·9066	·9173	·9257	·9326	·9383
	·95	·7370	·7919	·8514	·8839	·9045	·9189	·9295	·9376	·9440	·9493	·9536
	·99	·8498	·8826	·9173	·9358	·9475	·9556	·9615	·9660	·9695	·9724	·9748
9	0·80	0·4688	0·5526	0·6558	0·7189	0·7619	0·7933	0·8173	0·8363	0·8517	0·8644	0·8751
	·85	·5108	·5903	·6869	·7452	·7848	·8136	·8354	·8527	·8666	·8782	·8878
	·90	·5632	·6366	·7244	·7768	·8120	·8374	·8567	·8720	·8842	·8943	·9027
	·95	·6383	·7017	·7761	·8197	·8487	·8696	·8853	·8976	·9076	·9157	·9225
	·99	·7635	·8074	·8575	·8862	·9051	·9185	·9286	·9364	·9427	·9478	·9521
11	0·80	0·4003	0·4810	0·5859	0·6536	0·7016	0·7376	0·7657	0·7884	0·8069	0·8225	0·8357
	·85	·4389	·5169	·6169	·6808	·7257	·7592	·7854	·8063	·8235	·8378	·8500
	·90	·4880	·5617	·6551	·7138	·7548	·7854	·8089	·8278	·8433	·8561	·8670
	·95	·5603	·6267	·7091	·7600	·7952	·8212	·8413	·8573	·8702	·8810	·8902
	·99	·6878	·7381	·7989	·8357	·8607	·8790	·8929	·9039	·9128	·9202	·9265
13	0·80	0·3490	0·4253	0·5286	0·5982	0·6489	0·6880	0·7190	0·7443	0·7654	0·7832	0·7984
	·85	·3843	·4590	·5589	·6252	·6735	·7103	·7395	·7632	·7829	·7996	·8138
	·90	·4298	·5016	·5965	·6587	·7035	·7375	·7644	·7862	·8042	·8194	·8324
	·95	·4981	·5646	·6507	·7063	·7459	·7757	·7992	·8181	·8337	·8468	·8580
	·99	·6233	·6770	·7446	·7872	·8171	·8394	·8568	·8706	·8821	·8915	·8997
15	0·80	0·3091	0·3809	0·4812	0·5508	0·6030	0·6439	0·6769	0·7042	0·7271	0·7467	0·7636
	·85	·3415	·4124	·5102	·5775	·6275	·6664	·6978	·7236	·7453	·7638	·7798
	·90	·3837	·4527	·5468	·6106	·6577	·6942	·7235	·7475	·7675	·7847	·7993
	·95	·4478	·5130	·6003	·6584	·7011	·7338	·7598	·7810	·7989	·8138	·8268
	·99	·5687	·6237	·6954	·7422	·7758	·8013	·8216	·8378	·8512	·8629	·8726
17	0·80	0·2774	0·3447	0·4412	0·5101	0·5628	0·6047	0·6389	0·6676	0·6919	0·7129	0·7312
	·85	·3072	·3741	·4690	·5360	·5869	·6272	·6600	·6874	·7105	·7304	·7478
	·90	·3463	·4122	·5043	·5685	·6169	·6550	·6860	·7116	·7334	·7519	·7681
	·95	·4065	·4697	·5564	·6160	·6605	·6951	·7232	·7464	·7659	·7825	·7969
	·99	·5222	·5773	·6512	·7008	·7373	·7652	·7875	·8061	·8212	·8346	·8458
19	0·80	0·2515	0·3148	0·4073	0·4748	0·5273	0·5697	0·6047	0·6344	0·6597	0·6817	0·7010
	·85	·2791	·3424	·4338	·4999	·5509	·5919	·6257	·6541	·6785	·6995	·7179
	·90	·3155	·3782	·4677	·5315	·5805	·6196	·6517	·6786	·7016	·7214	·7387
	·95	·3719	·4327	·5182	·5782	·6238	·6599	·6894	·7139	·7349	·7528	·7684
	·99	·4823	·5369	·6116	·6630	·7014	·7313	·7555	·7756	·7926	·8071	·8198
21	0·80	0·2300	0·2895	0·3782	0·4439	0·4959	0·5383	0·5738	0·6040	0·6301	0·6529	0·6729
	·85	·2557	·3155	·4034	·4682	·5189	·5602	·5946	·6237	·6489	·6707	·6900
	·90	·2897	·3493	·4358	·4988	·5479	·5875	·6204	·6482	·6721	·6929	·7112
	·95	·3427	·4012	·4847	·5445	·5906	·6277	·6581	·6838	·7058	·7248	·7415
	·99	·4479	·5014	·5762	·6285	·6685	·6997	·7254	·7469	·7652	·7810	·7946

This table gives the values of x for which $\Pr(\theta_{\max.} < x) = I_x(2; p, q) = P$, where $p = \frac{1}{2}(\nu_2 - 1)$, $q = \frac{1}{2}(\nu_1 - 1)$.

Generalized Beta distribution (cont.)

ν_1 \ ν_2	P	2	3	5	7	9	11	13	15	17	19	21
23	0·80	0·2119	0·2680	0·3528	0·4167	0·4678	0·5101	0·5457	0·5763	0·6028	0·6262	0·6469
	·85	·2359	·2924	·3769	·4401	·4903	·5315	·5661	·5958	·6215	·6441	·6640
	·90	·2677	·3244	·4080	·4699	·5185	·5584	·5918	·6202	·6448	·6663	·6853
	·95	·3177	·3737	·4551	·5143	·5606	·5981	·6294	·6558	·6787	·6986	·7160
	·99	·4179	·4701	·5443	·5971	·6380	·6703	·6970	·7197	·7391	·7558	·7705
25	0·80	0·1964	0·2494	0·3306	0·3926	0·4428	0·4846	0·5202	0·5509	0·5777	0·6015	0·6226
	·85	·2189	·2725	·3536	·4151	·4645	·5055	·5402	·5702	·5963	·6193	·6397
	·90	·2488	·3027	·3834	·4439	·4921	·5319	·5655	·5944	·6194	·6415	·6610
	·95	·2960	·3498	·4287	·4872	·5333	·5710	·6027	·6298	·6533	·6738	·6920
	·99	·3915	·4424	·5155	·5685	·6096	·6429	·6708	·6941	·7143	·7319	·7474
27	0·80	0·1830	0·2333	0·3110	0·3711	0·4202	0·4615	0·4968	0·5275	0·5546	0·5785	0·6000
	·85	·2042	·2551	·3331	·3928	·4413	·4818	·5165	·5465	·5728	·5962	·6170
	·90	·2324	·2839	·3616	·4206	·4682	·5077	·5413	·5704	·5958	·6183	·6383
	·95	·2771	·3286	·4052	·4626	·5084	·5462	·5781	·6056	·6296	·6506	·6693
	·99	·3682	·4176	·4895	·5422	·5837	·6175	·6458	·6700	·6909	·7092	·7254
29	0·80	0·1713	0·2191	0·2936	0·3518	0·3998	0·4404	0·4754	0·5060	0·5331	0·5572	0·5789
	·85	·1913	·2398	·3147	·3727	·4202	·4603	·4947	·5247	·5512	·5747	·5958
	·90	·2180	·2672	·3420	·3996	·4463	·4855	·5191	·5482	·5738	·5966	·6170
	·95	·2604	·3099	·3840	·4404	·4855	·5232	·5553	·5830	·6074	·6288	·6480
	·99	·3475	·3954	·4659	·5182	·5597	·5938	·6225	·6472	·6687	·6876	·7043
31	0·80	0·1611	0·2065	0·2780	0·3344	0·3812	0·4211	0·4557	0·4861	0·5131	0·5373	0·5591
	·85	·1800	·2262	·2982	·3546	·4010	·4405	·4746	·5045	·5309	·5546	·5759
	·90	·2053	·2523	·3246	·3805	·4264	·4651	·4985	·5277	·5534	·5763	·5969
	·95	·2457	·2931	·3650	·4200	·4647	·5022	·5342	·5620	·5866	·6082	·6278
	·99	·3290	·3754	·4444	·4961	·5374	·5717	·6008	·6258	·6477	·6671	·6843
33	0·80	0·1519	0·1953	0·2639	0·3186	0·3643	0·4034	0·4375	0·4677	0·4946	0·5187	0·5406
	·85	·1699	·2141	·2834	·3381	·3835	·4223	·4560	·4857	·5121	·5358	·5572
	·90	·1940	·2389	·3087	·3632	·4081	·4464	·4794	·5084	·5342	·5572	·5781
	·95	·2324	·2781	·3478	·4015	·4455	·4825	·5145	·5424	·5671	·5890	·6088
	·99	·3123	·3573	·4247	·4757	·5168	·5511	·5803	·6057	·6279	·6476	·6653
35	0·80	0·1438	0·1852	0·2512	0·3042	0·3487	0·3871	0·4208	0·4506	0·4773	0·5014	0·5233
	·85	·1609	2032	·2699	·3230	·3674	·4055	·4388	·4682	·4945	·5181	·5396
	·90	·1838	·2270	·2943	·3473	·3913	·4290	·4617	·4906	·5163	·5394	·5603
	·95	·2206	·2645	·3320	·3845	·4277	·4644	·4962	·5240	·5487	·5708	·5907
	·99	·2972	·3408	·4066	·4569	·4974	·5318	·5612	·5867	·6091	·6292	·6471
37	0·80	0·1365	0·1762	0·2397	0·2910	0·3344	0·3721	0·4052	0·4347	0·4612	0·4852	0·5070
	·85	·1528	·1933	·2577	·3092	·3526	·3900	·4228	·4519	·4781	·5017	·5231
	·90	·1747	·2161	·2812	·3328	·3759	·4129	·4453	·4739	·4995	·5226	·5434
	·95	·2098	·2521	·3175	·3689	·4113	·4475	·4791	·5068	·5315	·5536	·5737
	·99	·2834	·3257	·3900	·4394	·4797	·5138	·5430	·5688	·5915	·6117	·6299
39	0·80	0·1299	0·1679	0·2292	0·2790	0·3213	0·3582	0·3907	0·4198	0·4460	0·4699	0·4916
	·85	·1455	·1844	·2465	·2965	·3389	·3756	·4079	·4367	·4627	·4861	·5076
	·90	·1664	·2062	·2691	·3194	·3616	·3980	·4299	·4583	·4837	·5067	·5277
	·95	·2001	·2408	·3044	·3544	·3961	·4319	·4631	·4906	·5152	·5375	·5576
	·99	·2709	·3119	·3747	·4232	·4631	·4969	·5260	·5517	·5747	·5950	·6134

This table gives the values of x for which $\Pr(\theta_{\max.} < x) = I_x(2; p, q) = P$, where $p = \frac{1}{2}(\nu_2 - 1)$, $q = \frac{1}{2}(\nu_1 - 1)$.

Generalized Beta distribution (cont.)

ν_1 \ ν_2	P	2	3	5	7	9	11	13	15	17	19	21
41	0·80	0·1239	0·1604	0·2195	0·2679	0·3091	0·3452	0·3772	0·4059	0·4319	0·4555	0·4771
	·85	·1388	·1762	·2362	·2849	·3262	·3622	·3940	·4225	·4482	·4715	·4928
	·90	·1589	·1972	·2582	·3070	·3483	·3840	·4155	·4436	·4689	·4918	·5127
	·95	·1912	·2306	·2922	·3411	·3819	·4172	·4480	·4754	·4999	·5221	·5423
	·99	·2594	·2993	·3605	·4079	·4472	·4812	·5100	·5358	·5585	·5792	·5977
51	0·80	0·1006	0·1311	0·1814	0·2233	0·2598	0·2923	0·3216	0·3482	0·3725	0·3950	0·4157
	·85	·1129	·1443	·1955	·2380	·2748	·3073	·3366	·3631	·3874	·4097	·4303
	·90	·1295	·1619	·2142	·2571	·2941	·3267	·3559	·3823	·4063	·4284	·4487
	·95	·1565	·1900	·2434	·2868	·3239	·3564	·3853	·4114	·4350	·4566	·4765
	·99	·2140	·2486	·3029	·3462	·3828	·4144	·4424	·4675	·4900	·5106	·5295
61	0·80	0·0846	0·1109	0·1545	0·1914	0·2240	0·2534	0·2801	0·3047	0·3274	0·3485	0·3682
	·85	·0951	·1221	·1667	·2043	·2372	·2668	·2936	·3182	·3409	·3620	·3816
	·90	·1093	·1373	·1830	·2211	·2544	·2842	·3111	·3357	·3583	·3792	·3987
	·95	·1324	·1615	·2086	·2474	·2810	·3109	·3378	·3622	·3847	·4054	·4246
	·99	·1821	·2126	·2610	·3004	·3341	·3637	·3903	·4142	·4361	·4561	·4743
71	0·80	0·0730	0·0960	0·1345	0·1675	0·1969	0·2236	0·2481	0·2708	0·2919	0·3117	0·3303
	·85	·0822	·1059	·1454	·1789	·2087	·2357	·2604	·2832	·3043	·3241	·3427
	·90	·0946	·1191	·1597	·1939	·2242	·2514	·2762	·2991	·3203	·3401	·3586
	·95	·1148	·1404	·1824	·2175	·2481	·2756	·3006	·3236	·3447	·3644	·3828
	·99	·1584	·1855	·2293	·2651	·2963	·3239	·3488	·3715	·3924	·4118	·4298
81	0·80	0·0643	0·0847	0·1191	0·1489	0·1756	0·2000	0·2226	0·2437	0·2634	0·2819	0·2994
	·85	·0723	·0934	·1288	·1592	·1863	·2110	·2338	·2550	·2748	·2934	·3109
	·90	·0833	·1052	·1417	·1727	·2003	·2253	·2483	·2697	·2896	·3082	·3257
	·95	·1013	·1242	·1620	·1939	·2221	·2475	·2707	·2922	·3122	·3308	·3483
	·99	·1402	·1646	·2044	·2374	·2662	·2919	·3153	·3368	·3567	·3751	·3924
91	0·80	0·0574	0·0757	0·1069	0·1340	0·1585	0·1810	0·2019	0·2215	0·2399	0·2573	0·2738
	·85	·0646	·0836	·1156	·1433	·1682	·1910	·2122	·2319	·2505	·2680	·2845
	·90	·0745	·0942	·1273	·1556	·1810	·2042	·2255	·2455	·2642	·2818	·2984
	·95	·0906	·1114	·1457	·1750	·2010	·2245	·2462	·2664	·2852	·3029	·3195
	·99	·1257	·1479	·1844	·2149	·2416	·2657	·2874	·3081	·3270	·3444	·3608
101	0·80	0·0518	0·0685	0·0969	0·1218	0·1444	0·1652	0·1847	0·2029	0·2202	0·2366	0·2522
	·85	·0584	·0756	·1049	·1304	·1533	·1745	·1942	·2127	·2301	·2466	·2622
	·90	·0673	·0853	·1155	·1416	·1651	·1866	·2066	·2253	·2428	·2595	·2752
	·95	·0820	·1009	·1325	·1594	·1835	·2055	·2258	·2447	·2625	·2792	·2951
	·99	·1140	·1343	·1678	·1961	·2211	·2436	·2644	·2836	·3015	·3184	·3342
121	0·80	0·0434	0·0575	0·0817	0·1030	0·1225	0·1407	0·1577	0·1739	0·1892	0·2038	0·2178
	·85	·0489	·0635	·0885	·1104	·1303	·1487	·1660	·1824	·1978	·2126	·2267
	·90	·0565	·0717	·0976	·1200	·1404	·1593	·1768	·1934	·2091	·2240	·2382
	·95	·0688	·0849	·1120	·1353	·1563	·1756	·1936	·2105	·2264	·2415	·2559
	·99	·0960	·1133	·1423	·1670	·1891	·2090	·2274	·2447	·2610	·2763	·2907
161	0·80	0·0328	0·0435	0·0622	0·0788	0·0941	0·1085	0·1221	0·1351	0·1476	0·1596	0·1711
	·85	·0370	·0481	·0674	·0844	·1001	·1148	·1287	·1419	·1545	·1666	·1783
	·90	·0427	·0544	·0744	·0920	·1081	·1231	·1372	·1507	·1635	·1758	·1876
	·95	·0521	·0645	·0856	·1039	·1206	·1361	·1506	·1644	·1775	·1900	·2020
	·99	·0730	·0864	·1092	·1288	·1464	·1627	·1778	·1921	·2056	·2185	·2309

This table gives the values of x for which $\Pr(\theta_{\max} < x) = I_x(2; p, q) = P$, where $p = \frac{1}{2}(\nu_2 - 1)$, $q = \frac{1}{2}(\nu_1 - 1$

MISCELLANEA

On problems in which a change in a parameter occurs at an unknown point

By E. S. PAGE
Department of Mathematics, University of Durham

1. Introduction

Given a sample of n independent observations in the order in which they were obtained, $x_1, \ldots, x_n$; we consider the problem of testing the null hypothesis that all the observations come from the same population with distribution function $F(x \mid \theta)$ against the alternatives that the first m $(0 \leqslant m < n)$, $x_1, \ldots, x_m$ come from $F(x \mid \theta)$ and $x_{m+1}, \ldots, x_n$ come from $F(x \mid \theta')$ $(\theta' \neq \theta)$, where m is unknown. In an earlier paper (Page, 1955) a one-sided test for a change in the mean of a distribution was proposed; the procedure was to record the cumulative sum $S_r = \sum_{j=1}^{r} (x_j - \theta)$, and, if the mean path after the change had a greater slope than that before the change, to use as a test statistic the rise in the cumulative sum above its least value, i.e. $\max_{0 \leqslant r \leqslant n} (S_r - \min_{0 \leqslant i < r} S_i)$, $S_0 = 0$, large values being significant. Some critical values were given for the case where the x_i are 0 or 1 binomial variables. Here, a discrimination approach is adopted for the more general case and the procedure is shown to yield a modification of the previous test.

2. One-sided case

Suppose θ, θ' known, and let the sample be $\mathbf{x} = (x_1, \ldots, x_n)$. If we regard the detection of a change as a problem in discrimination we have $n+1$ hypotheses between which it is desired to discriminate; they are $H_0, H_1, \ldots, H_n$, where H_i is the hypothesis that the first i observations are drawn from $F(x \mid \theta)$ and the remainder from $F(x \mid \theta')$, i.e. that the change occurs after the ith observation. A method of discrimination is specified by the definition of a division of the whole sample space $\mathfrak{S}$ into mutually exclusive regions R_i $(i = 0, 1, \ldots, n)$; the hypothesis H_i is preferred if the sample point falls within R_i. If the H_i are *a priori* equally likely the minimum probability of misclassification is obtained when the regions $R_i (i = 0, 1, \ldots, n)$ are defined by (Rao, 1948),

$$\mathbf{x} \in R_i \quad \text{if} \quad L_i(\mathbf{x}) \geqslant L_j(\mathbf{x}) \quad (j \neq i), \tag{2·1}$$

where $L_i(\mathbf{x})$ is the likelihood of the hypothesis H_i.

Example 1. *Mean of a normal distribution*

Suppose that we wish to discriminate for a change in the mean from μ to $\mu + \delta$ $(\delta > 0)$ with the standard deviation, σ, known and constant. Then (2·1) gives R_i defined by

$$\exp -\frac{1}{2\sigma^2} \left\{ \sum_{j=i+1}^{i+k} [(x_j - \mu - \delta)^2 - (x_j - \mu)^2] \right\} \geqslant 1 \quad (k = 1, \ldots, n-i),$$

$$\exp -\frac{1}{2\sigma^2} \sum_{j=i-k}^{i} \left\{ [(x_j - \mu - \delta)^2 - (x_j - \mu)^2] \right\} \leqslant 1 \quad (k = 0, 1, \ldots, n-i).$$

Hence R_i is defined by

$$\sum_{j=i+1}^{i+k} \left(x_j - \mu - \frac{\delta}{2} \right) \geqslant 0 \quad (k = 1, \ldots, n-i), \tag{2·2}$$

$$\sum_{j=i-k}^{i} \left(x_j - \mu - \frac{\delta}{2} \right) \leqslant 0 \quad (k = 0, 1, \ldots, i-1). \tag{2·3}$$

The discrimination procedure can be more clearly described in terms of the cumulative sum of the $(x - \mu - \frac{1}{2}\delta)$. Let $S_k = \sum_{j=1}^{k} (x_j - \mu - \frac{1}{2}\delta)$, $S_0 = 0$. Then the sample point $\mathbf{x} \equiv (x_1, \ldots, x_n)$ lies within R_i if $S_i \leqslant S_j$ for all j. For (2·2) is

$$\left.\begin{aligned} S_{i+k} - S_i &\geqslant 0 \quad (k = 1, \ldots, n-i), \\ \text{and (2·3) is} \quad S_i - S_{i-k-1} &\leqslant 0 \quad (k = 0, 1, \ldots, i-1). \end{aligned}\right\} \tag{2·4}$$

Hence S_i is the least of the cumulative sums. Thus the procedure may be conveniently carried out by recording the cumulative sum on a chart and selecting the hypothesis corresponding to the minimum of the sum.

Example 2. *Distributions with a sufficient statistic*

Consider distributions with frequency functions of Koopman's form

$$f(x \mid \theta) = \exp\{A(\theta)\,B(x) + C(\theta) + D(x)\}.$$

Then
$$f'(x)/f(x) = \exp\{B(x)\,\Delta A(\theta) + \Delta C(\theta)\},$$

where $\Delta G(\theta) = G(\theta') - G(\theta)$. Accordingly the regions R_i are given by

$$\left.\begin{aligned} \Delta A(\theta) \sum_{j=i+1}^{i+k} B(x_j) + k\Delta C(\theta) \geqslant 0 \quad (k = 1, \ldots, n-i), \\ \Delta A(\theta) \sum_{j=i-1}^{i} B(x_j) + (k+1)\,\Delta C(\theta) \leqslant 0 \quad (k = 0, \ldots, i). \end{aligned}\right\} \tag{2·5}$$

If $\Delta A(\theta) > 0$, the cumulative sum
$$S_i = \sum_{j=1}^{i} \left\{B(x_j) + \frac{\Delta C(\theta)}{\Delta A(\theta)}\right\} \tag{2·6}$$

may be recorded and its minimum used to select the hypothesis preferred. If $\Delta A(\theta) < 0$, the maximum of the same cumulative sum is used in the discrimination procedure.

Thus for a normal population whose mean remains constant at μ but whose variance may change from σ_1^2 to σ_2^2, the cumulative sum to be recorded is

$$S_i = \sum_{j=1}^{i} \left\{(x_i - \mu)^2 + 2\left[\frac{1}{\sigma_2^2} - \frac{1}{\sigma_1^2}\right]^{-1} \log\frac{\sigma_2}{\sigma_1}\right\}.$$

In some situations it may be *a priori* more probable that the initial value of the parameter will persist throughout the sampling than that a change should occur. If the hypothesis H_n, that no change from the initial value has occurred, is *a priori* c times as likely as any of the other hypotheses, then the regions R_i, for the discrimination procedure are defined by

$$\left.\begin{aligned} \mathbf{x} \in R_i \quad \text{if} \quad L_i(\mathbf{x}) \geqslant L_j(\mathbf{x}) \quad (j \neq i,\ i, j \neq n), \\ \geqslant cL_n(\mathbf{x}) \quad (i \neq n), \\ \text{and} \qquad \mathbf{x} \in R_n \quad \text{if} \quad cL_n(\mathbf{x}) \geqslant L_i(\mathbf{x}) \quad (i \neq n). \end{aligned}\right\} \tag{2·7}$$

For distributions with a sufficient statistic the above conditions amend (2·5) only for the comparison between H_i and H_n. H_n is preferred to H_i if

$$\Delta A(\theta) \sum_{j=i+1}^{n} B(x_j) + (n-i)\,\Delta C(\theta) \geqslant \log c. \tag{2·8}$$

Accordingly the hypothesis to be preferred is that indicated by the minimum (for $\Delta A(\theta) > 0$) of the cumulative sum, S_j, of (2·6) if the rise of S_n above its least value exceeds $\log c/\Delta A(\theta)$; otherwise the hypothesis that no change has occurred is adopted. For the test on a normal mean with unit variance this rise in its minimum is $\delta^{-1}\log c$.

3. A ONE-SIDED TEST

The procedure of § 2 is one that can reasonably be used as a test of the hypothesis that the observations $x_1, \ldots, x_n$ are all drawn from the same population, $F(x \mid \theta)$. If $F(x \mid \theta)$ is of Koopman's form the appropriate cumulative sum is plotted and the null hypothesis rejected if the final point of the sample path is at least a given distance above its minimum ($\Delta A > 0$) (Fig. 1). We note that in this test the criterion is the distance of the end-point of the path above the minimum value, whereas in the earlier test (Page, 1955) we considered the greatest distance of the path above the minimum. The properties of this test are likely to be difficult to calculate as it is equivalent to a truncated sequential test. However, an approximation to the power function which is valid for a large sample and when a large rise in the cumulative sum is required for the rejection of H_n may be obtained from the normal diffusion process in the presence of an absorbing barrier. Since the test is a rise in the cumulative sum at the end of the path, we may express it in terms of the end-point as a fall below that value. Thus if we look at the sample in the reverse order, the criterion is whether or not the cumulative sum crosses a fixed horizontal boundary within n steps; that is, the problem is that of a truncated random walk in the presence of an absorbing barrier.

A very simple approximation to the probability that the boundary is crossed within n steps when the increments, y_i, have some symmetrical distribution about zero mean and standard deviation σ, has been given by Armitage (1957). Let p_i be the probability that the cumulative sum Σy_j exceeds a constant

k for the first time after the ith observation. Then the probability that the final sum exceeds k given that the sum first did so after the ith observation is slightly greater than $\frac{1}{2}p_i$. Hence by summation,

$$\begin{aligned}\Pr\{S_n \geqslant k\} &\dot{>} \frac{1}{2}\sum_{i=1}^{n-1} p_i + p_n \\ &\dot{>} \frac{1}{2}\sum_{i=1}^{n} p_i \\ &= \tfrac{1}{2}\Pr(\text{boundary is crossed within } n \text{ steps}),\end{aligned}$$

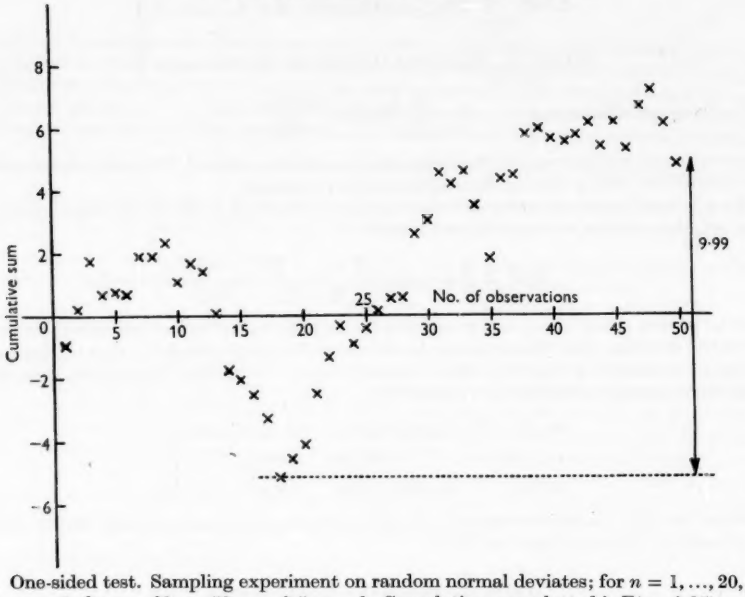


Fig. 1. One-sided test. Sampling experiment on random normal deviates; for $n = 1, \ldots, 20, \mu = 0$, $\sigma = 1$; for $n = 21, \ldots, 50, \mu = 0{\cdot}5, \sigma = 1$. Cumulative sum plotted is $\Sigma(x - 0{\cdot}25)$.

where the symbol $\dot{>}$ means 'slightly greater-than'. Hence

$$\Pr(\text{boundary is crossed}) \doteqdot 2\left[1 - \Phi\left(\frac{k}{\sigma\sqrt{n}}\right)\right], \tag{3·1}$$

using the central limit theorem, where $\Phi(t) = (2\pi)^{-\frac{1}{2}}\int_{-\infty}^{t} \exp(-\frac{1}{2}x^2)\,dx$.

A more general approximation for increments with non-zero mean, which we need here, is obtained as in Cox's (unpublished) treatment of Armitage's problem, by replacing the random walk in discrete time by one in continuous time. For such a process in which the increment per unit time has mean m and variance σ^2, the probability density of the cumulative sum at time t is

$$f(y, b, t) = (2\pi\sigma^2 t)^{-\frac{1}{2}}\left[\exp -\frac{1}{2\sigma^2}\left(\frac{y}{\sqrt{t}} - m\sqrt{t}\right)^2 - \exp\frac{2bm}{\sigma^2}\exp -\frac{1}{2\sigma^2}\left(\frac{y-2b}{\sqrt{t}} - m\sqrt{t}\right)^2\right], \tag{3·2}$$

where the boundary is at $y = b$ ($b > 0$) (e.g. Bartlett, 1955, p. 48). The probability that the boundary has not been crossed in the interval (O, T) is therefore

$$g(b, T) = \int_{-\infty}^{b} f(y, b, T)\,dy = \Phi\left(\frac{b}{\sigma\sqrt{T}} - \frac{m\sqrt{T}}{\sigma}\right) - e^{2mb/\sigma^2}\Phi\left(-\frac{b}{\sigma\sqrt{T}} - \frac{m\sqrt{T}}{\sigma}\right). \tag{3·3}$$

Hence given m and σ^2 the mean and variance of the increment, y, and $T = n$, the number of observations, we can apply this result directly to our test to find the position, b, of the boundary so that the hypothesis that all the x's come from the specified distribution will be rejected when it is true with probability approxi-

mately any given value. A few values are shown in Table 1 for a probability 0·05 of Type I errors. For example, to test whether in fifty observations a change from μ to $\mu+0{\cdot}2\sigma$ occurs in the mean of a normal population, the sum $\Sigma(x_i-\mu-0{\cdot}1\sigma)$ is plotted in the presence of a boundary at $10{\cdot}0\sigma$.

Table 1. *Approximate values of* b/σ

m/σ \ n	25	50	100
−0·025	9·3	12·8	17·6
−0·05	8·8	11·8	15·6
−0·10	7·8	10·0	12·2
−0·20	6·1	6·9	7·4

The diffusion approximation is of the same kind as the Wald approximation for the characteristics of sequential tests and, as the latter are usually sufficient for practical purposes, it is reasonable to suppose that in this case the same holds. By an enumeration of paths to the boundary, it is possible to evaluate exactly the probabilities for the binomial case and these agree fairly well with those given by the approximate formulae (Armitage, 1957).

4. Two-sided case

Let us now consider the case where the alternatives to the hypothesis that the observations all come from a distribution $F(x\,|\,\theta)$ are that the first i observations come from $F(x\,|\,\theta)$ and the remainder all come from either $F(x\,|\,\theta')$ or $F(x\,|\,\theta'')$, where $\theta'>\theta>\theta''$, i.e. the change in the parameter may be in either direction. We suppose θ, θ', θ'' known. Then in the discrimination problem we have $2n+1$ completely specified hypotheses H_n, H_i^+, H_i^- $(i=0,1,\ldots,n-1)$, where H_i^+ (H_i^-) are the hypotheses that the first i observations come from $F(x\,|\,\theta)$ and the remainder from $F(x\,|\,\theta')$, $(F(x\,|\,\theta''))$; the superfix $+$, $-$ thus indicates the direction of the change in θ. As before, we divide the whole sample space into $(2n+1)$ mutually exclusive regions R_n, R_i^+, R_i^- $(i=0,\ldots,n-1)$ such that the region in which the sample point falls indicates the hypothesis preferred. When the hypotheses are *a priori* equally likely the probability of misclassification (or the expected loss if the weighted losses are equal) is minimized if the regions are defined by:

$$\left.\begin{aligned}&\mathbf{x}\in R_i^+ \quad \text{if} \quad L_i^+(\mathbf{x})\geqslant\max\left(L_j^+(\mathbf{x}),L_k^-(\mathbf{x}),L_n(\mathbf{x})\right) \quad (0\leqslant j;\ k\leqslant n-1;\ j\neq i;\ i=0,1,\ldots,n-1).\\ &\mathbf{x}\in R_n \quad \text{if} \quad L_n(\mathbf{x})\geqslant\max\left(L_i^+(\mathbf{x}),L_i^-(\mathbf{x})\right) \quad (0\leqslant i\leqslant n-1),\end{aligned}\right\} \tag{4·1}$$

with a similar definition for R_i^-, where $L_i^+(\mathbf{x})$, $L_i^-(\mathbf{x})$, $L_n(\mathbf{x})$ are the likelihoods of the corresponding hypotheses. A comparison of the H^+'s (H^-'s) with H_n leads, of course, to the one-sided criterion for discrimination. If H_n is preferred under both the one-sided procedures, it is preferred under the two-sided procedure; if H_n is preferred by one, but a different hypothesis, say H_i^*, by the other, then H_i^* is preferred by the two-sided procedure. If each one-sided procedure prefers some hypothesis other than H_n it is necessary to use the additional conditions implied in (3·1) to choose which of the two is preferred by the two-sided procedure.

Example 3. *Mean of a normal distribution*

Suppose that the three possible means are 0, $\pm\mu$, and that the variance is unity. Then H_n, i.e. that no change has occurred, is preferred if

$$\left.\begin{aligned}&\max_{0\leqslant i\leqslant n-1}\sum_{k=i+1}^{n}(x_k-\tfrac{1}{2}\mu)\leqslant 0,\\ \text{and}\quad &\min_{0\leqslant j\leqslant n-1}\sum_{k=j+1}^{n}(x_k+\tfrac{1}{2}\mu)\geqslant 0.\end{aligned}\right\} \tag{4·2}$$

If just one of these inequalities is unsatisfied, suppose it is the first and that the maximum occurs for $i=l$. Then H_l^+ is preferred. If both inequalities are unsatisfied, the maximum and minimum occurring at $i=l$, $j=m$, then H_l^+ is preferred to H_m^- if

$$\sum_{k=l+1}^{n}(x_k-\tfrac{1}{2}\mu)\geqslant-\sum_{k=m+1}^{n}(x_k+\tfrac{1}{2}\mu), \tag{4·3}$$

and conversely.

For a graphical representation it is convenient to consider the graph of the cumulative sum of the observations themselves, $S_j = \sum_{i=1}^{j} x_i$ against j. The conditions (4·2) and (4·3) are then expressible as whether or not the lines $y - S_n = \pm \frac{1}{2}\mu(x-n)$ contain the path and, if the path intersects both lines, which of the intersections is the deeper. The lines concerned are those drawn through the end point of the path, with slopes $\pm \frac{1}{2}\mu$.

More generally, for distributions of Koopman's form with a sufficient statistic and for which $A(\theta)$ is monotonic, the cumulative sum $\sum_{j=1}^{i} B(x_j)$ can be plotted and lines with slopes

$$s' = \{C(\theta') - C(\theta)\}/\{A(\theta') - A(\theta)\}, \quad s'' = \{C(\theta'') - C(\theta')\}/\{A(\theta'') - A(\theta')\}$$

drawn through the end point. It may be noted that when the procedure is applied as it stands to a sequence of 0, 1 values from a binomial population then H_n, the hypothesis that no change has occurred, is never preferred; if the last observation is '1', one of the H^+ hypotheses is preferred to H_n, and similarly if $x_n = 0$. This disadvantage does not necessarily persist if the observations are grouped.

If the hypothesis of no change, H_n, is *a priori* c times as likely as any of the alternative hypotheses the procedure gives rise to two sets of inequalities like (2·11) for the comparisons between H_n and the H^+'s or the H^-'s; in the case of Koopman's distributions these yield boundaries with slopes s', s'', which pass through the points $(n, S_n + d(\theta, \theta''))$, $(n, S_n + d(\theta, \theta'))$, where $d(\theta, \theta^*)$ is a function of θ, θ^* determined by the distribution of x. For comparisons of one of the H^+'s with an H^-, the condition is the same as when $c = 1$, i.e. it depends upon the deeper of the intersections with the above boundaries with $d \equiv 0$ (if $d(\theta, \theta'') = -d(\theta, \theta')$ this is of course equivalent to preferring the hypothesis indicated by the deeper intersection of the boundaries as drawn). An approximation to the positions of the boundaries if this procedure is to be used as a test can be obtained from the one-sided test and its diffusion process approximation. Armitage (1957) has shown that the error in this approximation is small enough to be neglected in most practical cases.

5. Conclusion and summary

A one-sided test for a change in a parameter within a set of observations in the order drawn is given, and an approximation to the boundary is derived. The test is shown to arise from a discrimination procedure and the corresponding two-sided test is quoted. These tests are similar to the 'restricted sequential procedures' proposed by Armitage in another context, where divergent boundaries and an upper limit to the sample size are imposed at the outset. Here our sample size is fixed and the boundaries are drawn at the end of the sampling.

It is interesting also to compare the procedures given here with those derived by Rao (1950) for sequential tests of null hypotheses when nothing is specified about the alternative hypotheses. Rao's approach used the locally most powerful best test to find a criterion which, if satisfied in less than a fixed number of observations, would lead to the rejection of the null hypothesis. Thus for a one-sided test of the mean, μ, of a normal population of known standard deviation, Rao's test is to reject the null hypothesis if $\sum_{i=1}^{n} (x_i - \mu) \geqslant k$ for some i $(0 \leqslant i \leqslant N)$, which is similar to our test when it is looked at in the reverse direction, although the cumulative sum plotted is different.

Rao's test for distributions of Koopman's form considers the cumulative sums of $B(x)A'(\theta) + C'(\theta)$, instead of using the differences as in (2·9), and will, of course, have a criterion different from (2·9) in general. For example, Rao's test for the standard deviation of a normal population of known mean plots a cumulative sum which has a mean path of slope zero on the null hypothesis while the cumulative sum of our procedure has mean path with slope different from zero on all the hypotheses.

I wish to thank Dr P. Armitage for reading this note in draft and for allowing me to see his paper on Restricted Sequential Procedures before publication.

REFERENCES

Armitage, P. (1957). *Biometrika*, **44**, 9.
Bartlett, M. S. (1955). *Stochastic Processes*. Cambridge University Press.
Lindley, D. V. (1953). *J. R. Statist. Soc.* B, **15**, 30.
Page, E. S. (1955). *Biometrika*, **42**, 523.
Rao, C. R. (1948). *J. R. Statist. Soc.* B, **10**, 159.
Rao, C. R. (1950). *Sankyhā*, **10**, 361.

Testing for departure from the exponential distribution*

By D. J. BARTHOLOMEW
Scientific Department, National Coal Board, London

1. Introduction

Occasions arise in statistical practice when it is necessary to test whether a random variable, t, has the exponential distribution

$$p(t) = \lambda e^{-\lambda t} \quad (t \geqslant 0),$$

on the basis of an observed sample containing n independent observations, say, $t_1, t_2, \ldots, t_n$. Examples are to be found in life testing and the study of the distribution of intervals between events occurring in time or space (see, for example, Maguire, Pearson & Wynn, 1952). The object of this paper is to compare the power of three tests of whether a sample is exponentially distributed. The statistics considered are:

$$M = -2\left\{\sum_{i=1}^{n} \log_e t_i - n \log \bar{t}\right\},$$

$$S = \sum_{i=1}^{n} t_i \Big/ \left(\sum_{i=1}^{n} t_i\right)^2,$$

$$\varpi = \sum_{i=1}^{n} |t_i - \bar{t}|/(2n\bar{t}),$$

where

$$\bar{t} = \frac{1}{n}\sum_{i=1}^{n} t_i.$$

These criteria have been discussed in the work of a number of writers, in particular by Darling (1953), who gave a unified treatment of the distribution theory.

The power function of a test can only be found by reference to some alternative to the null hypothesis. Which alternative is appropriate in a given circumstance depends entirely on the particular nature of the problem in hand. In this paper, therefore, our aim will be the general one of trying to answer such a question as, 'Under what kind of alternatives is S more (or less) powerful than ϖ?' Such information should be of value in deciding which test to use.

Four alternatives will be considered and the relative power of the three tests determined with respect to each of them. It has not been possible to obtain the exact power functions or even good approximations to them in the majority of cases. Instead, use has been made of the well-known result concerning the asymptotic relative efficiency of tests. An instructive discussion of this method has been given by Cox & Stuart (1955). In one case (see § 4 below) where a fair approximation to the power of both M and S has been obtained, the asymptotic method is shown to be satisfactory for sample sizes of the order of 100.

2. Alternatives to the exponential

The four alternatives with which we shall be concerned have been selected because they show the strong and weak points of the tests. For a full discussion of the ways in which these and other alternatives can arise reference may be made to Bartholomew (1955).

Alternative I

$$p(t) = \lambda\{\Gamma(a)\}^{-1}(\lambda t)^{a-1} e^{-\lambda t} \quad (t \geqslant 0,\ a > 0).$$

Moran (1951) suggested this alternative and showed that M was the asymptotically most powerful test of the hypothesis $a = 1$. The curve has a zero ordinate at $t = 0$ for $a > 1$ and an infinite ordinate if $0 < a < 1$.

Alternative II

$$p(t) = \lambda(\lambda t)^{a-1} \exp\{-(\lambda t)^a/a\} \quad (t \geqslant 0,\ a > 0).$$

This distribution is usually associated with the name of Weibull. It is similar in shape to I above for a near to unity, having initial values equal to infinity and zero for $0 < a < 1$ and $a > 1$ respectively.

Alternative III

$$p(t) = \lambda(1 + a\lambda t)^{-(a+1)/a} \quad (t \geqslant 0,\ a > 0).$$

In contrast to the preceding curves, this alternative is always J-shaped with a finite ordinate at $t = 0$. Maguire *et al.* gave it as an alternative to the exponential. It is a special case of a class considered by

* The work contained in this paper is based on part of a thesis approved for the Ph.D. degree in the University of London.

Neyman (1941) in connexion with testing for homogeneity of variance. He arrived at the criterion S as being appropriate for testing the hypothesis $a = 0$. In the Pearson system this is a Type XI curve.

Alternative IV
$$p(t) = \lambda e^{-\lambda t}\{1 + a(\tfrac{1}{2}\lambda^2 t^2 - 2\lambda t + 1)\} \quad (t \geqslant 0).$$

This curve is similar in shape to III. It is a special case of the Laguerre series which is obtained from the Pearson Type III distribution in the same way as the Gram-Charlier curves are obtained from the normal distribution. Cox & Smith (1952) gave it as a limiting form of the distribution of intervals between the events in a 'pooled output'.

3. Power of the tests

The analysis of this section is based on the following result quoted from Cox & Stuart (1955).

If there are two consistent test statistics ϕ_1 and ϕ_2, of a hypothesis H_0: $\theta = \theta_0$, the asymptotic relative efficiency (A.R.E.) is the reciprocal of the ratio of sample sizes required to attain the same power against the same alternative H_1, taking the limit as the sample sizes tend to infinity and as H_1 tends to H_0.

If ϕ_1 and ϕ_2 both have normal limiting distributions on H_0 and H_1, the A.R.E. of ϕ_1 compared to ϕ_2 is given by
$$\text{A.R.E.}\,(\phi_1, \phi_2) = \lim \left(\frac{R^2(\phi_1)}{R^2(\phi_2)}\right)^{1/r},$$

where
$$R^2(\phi_i) = \left\{\frac{\partial}{\partial\theta}\mathscr{E}(\phi_i \mid \theta)\Big|_{\theta=\theta_0}\right\}^2 \Big/ \operatorname{var}(\phi_i \mid \theta = \theta_0),$$

provided that r satisfies the equations
$$\lim_{n\to\infty} R^2(\phi_i)\, n^{-r} = R_i \quad (i = 1, 2).$$

The R_i are constants independent of n which is the sample size. In all cases considered in this paper $r = 1$. For brevity we shall write
$$E = \frac{\partial}{\partial\theta}\mathscr{E}(\phi \mid \theta)\Big|_{\theta=\theta_0}, \quad V = \operatorname{var}(\phi \mid \theta = \theta_0).$$

It should be noted that it is only necessary to know the limiting values of E and V in order to calculate the A.R.E. This fact will enable us to find $\lim R^2$ when the mean value of the criterion under the alternative cannot be found exactly.

As to the condition of limiting normality it is known that each of the three statistics satisfy the condition on the null hypothesis. It is known further from a theorem of Moran (1947) that S satisfies it under any alternative. No such general result appears to be available for M and ϖ although in the special case of the Type III alternative M is normally distributed in the limit. It seems very likely that M and ϖ are asymptotically distributed in the normal form in the limit but even if this is not so the A.R.E. still provides a useful local comparison of power. The following results must be interpreted in the light of this remark.

We now proceed to calculate E for M, S and ϖ under the four alternatives listed above. For alternatives I and II the null hypothesis corresponds to $a = 1$ and for III and IV to $a = 0$.

I. (*a*) When the observations are distributed in the type III form it can be readily shown that
$$\mathscr{E}(M) = -2\{n\log n - n\psi(a) + n\psi(na)\},$$

where
$$\psi(x) = \frac{d}{dx}\log_e \Gamma(x).$$

Differentiating we find
$$\frac{\partial\mathscr{E}(M)}{\partial a}\Big|_{a=1} = -2\{n\psi^{(1)}(1) - n^2\psi^{(1)}(n)\},$$

which, taking the limit as n tends to infinity, gives
$$\lim E = -2n(\psi^{(1)}(1) - 1). \tag{1}$$

We write 'lim' for $\lim_{n\to\infty}$ throughout this section.

(*b*) The expected value of S under this alternative was shown by Moran (1951) to be $(a+1)/(an+1)$. This leads immediately to the result
$$E = n^{-1}. \tag{2}$$

(*c*) The exact value of $\mathscr{E}(\varpi \mid a)$ has not been obtained but we have

$$\lim \mathscr{E}(\varpi \mid a) = \tfrac{1}{2}(\text{Type III mean deviation})/(\text{Type III mean}).$$

Straightforward calculation shows this to be

$$I_a(a) - I_a(a+1),$$

where

$$I_x(a) = \int_0^x t^{a-1} e^{-t} dt / \Gamma(a).$$

Differentiating and simplifying, we ultimately find

$$\lim E = \gamma/e + 1 - 2/e - 0{\cdot}6321, \tag{3}$$

where γ = Euler's constant = 0·5772 ....

II. (*a*) For the Weibull distribution we find

$$\lim \mathscr{E}(M) = 2n\{\log_e \Gamma(1+1/a) - \psi(1)/a\}$$

and

$$\lim E = 2n(\psi(1) - \psi(2)) = -2n. \tag{4}$$

(*b*)

$$\lim \mathscr{E}(S) = \Gamma(1+2/a)/n(\Gamma(1+1/a))^2$$

and

$$\lim E = 4(\psi(2) - \psi(3))/n = n^{-1}. \tag{5}$$

(*c*) The limiting value of the mean of ϖ takes the rather complicated form

$$\lim \mathscr{E}(\varpi) = \left(1 - e^{-c} + \sum_{r=0}^{\infty} \frac{(-1)^r}{r!} \frac{c^{r+1/a}}{r+1+1/a}\right) \Big/ (ac)^{1/a},$$

where

$$c = (\Gamma(1+1/a))^a;$$

thus

$$\lim E = -\left\{\psi(2)(1-2/e) + \sum_{r=0}^{\infty} \frac{(-1)^r}{r!} \frac{1}{(r+2)^2}\right\}.$$

Summing the series on the right numerically we reach the result

$$\lim E = -\{\psi(2)(1-2/e) + 0{\cdot}1645\}. \tag{6}$$

III. (*a*) Under the third alternative

$$\lim \mathscr{E}(M) = 2n\{a(1+\psi(2) - \psi(3)) - \psi(1)\},$$

and therefore

$$\lim E = 2n(1+\psi(2) - \psi(3)) = n. \tag{7}$$

(*b*)

$$\lim \mathscr{E}(S) = 2(1-a)/(1-2a),$$

$$\lim E = 2. \tag{8}$$

(*c*)

$$\lim \mathscr{E}(\varpi) = (1-a)^{1/a},$$

$$\lim E = \tfrac{1}{2}e. \tag{9}$$

IV. (*a*) Under the fourth alternative we find

$$\lim \mathscr{E}(M) = -\psi(1) - \tfrac{1}{2}a,$$

$$\lim E = -\tfrac{1}{2}. \tag{10}$$

(*b*)

$$\lim \mathscr{E}(S) = (1+2a)/n,$$

$$\lim E = 2. \tag{11}$$

(*c*)

$$\lim \mathscr{E}(\varpi) = (1 - \tfrac{1}{2}a)/e,$$

$$\lim E = -\tfrac{1}{2}e. \tag{12}$$

The foregoing results are brought together in Table 1. For each alternative the criteria are compared with the best test to give the A.R.E. The roman numerals at the head of the columns refer to the numbering of the alternatives in the text.

4. The reliability of the A.R.E.

The results of § 3 are asymptotic properties of the tests. The object of this section is to see how far these limiting properties hold good in finite samples for one special case. Approximations to the power of both M and S can be obtained under the Type III alternative by fitting curves having the same moments. The calculation of β_1 and β_2 for M under this alternative suggests using an approximation of the form

$$M \simeq C\chi_\nu^2,$$

where C and ν are determined from the first two moments of M. Investigations into the form of the distribution of S, in the null case (Bartholomew, 1955), suggest that it might be fairly well approximated by a lognormal distribution.

According to the figures in Table 1 we should expect the power of S using 100 observations to be almost the same as that of M on 39 observations. The power functions obtained in the way described above are compared in Table 2. In view of the approximate nature of the comparison the agreement is good; this suggests that the asymptotic properties may be used with some confidence for sample sizes of this order.

Table 1. *Values of the asymptotic relative efficiency of M, S and ϖ under four alternatives*

Test statistic	Limiting variance on null hypothesis	Alternative law			
		I	II	III	IV
M	$4n(\psi^{(1)}(1)-1)$	1·00	1·00	0·38	0·38
S	$4n^{-3}$	0·39	0·64	1·00	1·00
ϖ	$0{\cdot}0591n^{-1}$	0·63	0·83	0·57	0·57

Table 2. *Comparison of the power of M and S under the Type III alternative*

a	1·0	0·9	0·8	0·7	0·6	0·5	0·4
M, (39 obs.)	0·050	0·142	0·326	0·591	0·840	0·969	0·998
S, (100 obs.)	0·050	0·140	0·316	0·560	0·794	0·939	0·990

5. Concluding remarks

Table 1 shows that the form of the alternative has a pronounced effect on the relative power of the three tests considered. On alternatives I and II, M is clearly the best test to use while the very reverse is true for III and IV. The reason for this can be appreciated by noting that it is the relatively small intervals that are most important in determining whether M is significant. We should therefore expect M to be the most powerful under alternatives showing a marked departure from the exponential near $t = 0$. This is in fact the case for I and II which take either zero or infinite values at this point. On the other hand, the value of S, which is a sum of squares, is influenced most by the largest values and so would be sensitive to departure from the exponential at the upper tail. The ϖ test takes up an intermediate position in both cases, giving less weight to the small values than does M and more than S, and vice versa for the large values.

In some instances it may be possible to decide, *a priori*, on an alternative distribution, but it is doubtful whether this will be so in general. Under the latter circumstance ϖ could be used on the grounds that it will always be second best, whereas either of M or S could be the worst. In this way we minimize the maximum loss in power which might be incurred by a wrong choice of test.

Even if it is known that M provides the best test, it can only be used in practice if the small values of t are recorded to high accuracy. This is not often the case, in fact very small values are frequently recorded as zero, thus making M infinite. Bearing in mind that the significance of M can depend to a large extent on inaccuracies in recording, it would be safer to use ϖ when there is any doubt.

My thanks are due to Dr N. L. Johnson, for his advice throughout this work, and to the Department of Scientific and Industrial Research for the award of a maintenance grant during the period when the research was carried out.

REFERENCES

BARTHOLOMEW, D. J. (1955). Ph.D. Thesis, University of London.
COX, D. R. & SMITH, W. L. (1952). *Biometrika*, **40**, 1.
COX, D. R. & STUART, A. (1955). *Biometrika*, **42**, 80.
DARLING, D. A. (1953). *Ann. Math. Statist.* **24**, 239.
MAGUIRE, B. A., PEARSON, E. S. & WYNN, A. H. A. (1952). *Biometrika*, **39**, 168.
MORAN, P. A. P. (1947). *J. R. Statist. Soc. Suppl.* **9**, 92.
MORAN, P. A. P. (1951). *J. R. Statist. Soc.* B, **13**, 147.
NEYMAN, J. (1941). *Ann. Math. Statist.* **12**, 46.

The distribution of range in normal samples with $n = 200$

BY B. I. HARLEY AND E. S. PEARSON
University College, London

1. INTRODUCTORY

In his paper of 1925, Tippett considered the distribution of the range, w, in samples from a normal population, chiefly from the point of view of the moments, over the whole stretch from samples of $n = 2$ to $n = 1000$. He tabled the expected value of the range (in terms of the population standard deviation, σ, as unit) to five decimal places for $n = 2(1)1000$, but only gave values of the standard deviation and beta coefficients for $n = 2, 10, 20, 60, 100, 200, 500, 1000$. His standard deviations were given to three decimal places, but he considered that little reliance could be placed on the third decimal place in the β_1 and β_2 values. After 1925, attention was focused on the distribution of range in small samples, largely because of the usefulness of this statistic as a measure of variability in industrial quality control. Thus Pearson & Hartley's (1942) tables gave the probability integral of w to four decimal places for $n = 2(1)20$, by intervals of 0·05 in w.

Recently, however, suggestions have been made* that a comparison of the range and root mean-square estimators of the population σ may serve a useful purpose in quite large samples, either as a test of homogeneity or as a routine check on accuracy in computation. Tukey (1955) discussed certain methods of interpolation which he showed could be used with some success to calculate such quantities as the percentage points of w or of w/s† within the very open grid of large sample information which has been scarcely extended since Tippett's original calculations.‡ He pointed out, however, that satisfactory results could hardly be established finally until one or more highly accurate computations had been carried out between, say, $n = 100$ and 1000. The tables and derived values given below for the single case $n = 200$ are presented as a contribution towards this objective.

2. DESCRIPTION OF RESULTS

Table 1 gives the probability integral of the range, w, to seven decimal places at argument intervals of 0·25. Each value was computed independently by quadrature as described in § 3 below. The last figure may be in error by one or two units in the middle of the table, but it was decided to put the seven-figure results on record, rather than to cut down to six decimals. Table 2 giving the probability integral to four decimal places at the more convenient argument interval of 0·05, was derived by interpolation from the basic figures of Table 1. From the latter it was also possible to calculate the moments of the distribution of w by quadrature, using the formula

$$\mu'_r(w) = (-w_0)^r + r\int_0^\infty (W - w_0)^{r-1} Q(W)\, dW, \qquad (1)$$

* See for example, David, Hartley & Pearson (1954).

† Here s is the usual standard deviation estimate of σ calculated from the same sample as w.

‡ Additional values of the second moment of w at $n = 30, 45$ and 75 were computed for one of us (Pearson, 1932). With Ruben's (1954, p. 224) highly accurate values of the mean range now available for $n = 30, 45$, it is possible to improve on the figures given in 1932, as follows:

Sample size	30	45	75
Variance of w	0·479 783	0·435 684	0·389 01
Standard deviation	0·692 66	0·660 06	0·623 7

where w_0 is a convenient working origin near the mean, taken in this case as 5·5, and $P(W) = 1-Q(W)$ is the probability integral. As usual in carrying out a quadrature of this kind, when we wish to stretch available information to the limit, it is a little difficult to be sure of the exact degree of accuracy of the results obtained. As a check, we computed moments by applying quadrature to equation (1) in three ways: using (*a*) the seven decimal and (*b*) rounded off six decimal place values of $Q(W)$ at intervals of 0·25; (*c*) using only the alternate seven decimal place values, i.e. values at an argument interval of 0·50.

Table 1. *Basic values of the probability integral*

W	$P(W)$	W	$P(W)$	W	$P(W)$
		6·00	0·822 4992	9·00	0·999 9962
3·25	0·000 0000	6·25	·902 9409	9·25	·999 9988
3·50	·000 0003	6·50	·950 5744	9·50	·999 9996
3·75	·000 0181	6·75	·976 3683	9·75	1·000 0000
4·00	0·000 3917	7·00	0·989 3214		
4·25	·004 0133	7·25	·995 4158		
4·50	·022 5924	7·50	·998 1226		
4·75	·079 3777	7·75	·999 2642		
5·00	0·193 6026	8·00	0·999 7233		
5·25	·357 9398	8·25	·999 8999		
5·50	·539 0411	8·50	·999 9652		
5·75	·700 6038	8·75	·999 9883		

Table 2. *Derived table with smaller argument interval*

W	$P(W)$	W	$P(W)$	W	$P(W)$	W	$P(W)$	W	$P(W)$
3·75	0·0000	4·75	0·0794	5·75	0·7006	6·75	0·9764	7·75	0·9993
3·80	·0000	4·80	·0974	5·80	·7284	6·80	·9798	7·80	·9994
3·85	·0001	4·85	·1178	5·85	·7546	6·85	·9827	7·85	·9995
3·90	·0001	4·90	·1407	5·90	·7790	6·90	·9852	7·90	·9996
3·95	·0002	4·95	·1660	5·95	·8016	6·95	·9874	7·95	·9997
4·00	0·0004	5·00	0·1936	6·00	0·8225	7·00	0·9893	8·00	0·9997
4·05	·0007	5·05	·2233	6·05	·8417	7·05	·9909	8·05	·9998
4·10	·0011	5·10	·2548	6·10	·8593	7·10	·9923	8·10	·9998
4·15	·0017	5·15	·2880	6·15	·8753	7·15	·9935	8·15	·9998
4·20	·0027	5·20	·3225	6·20	·8898	7·20	·9946	8·20	·9999
4·25	0·0040	5·25	0·3579	6·25	0·9029	7·25	0·9954	8·25	0·9999
4·30	·0059	5·30	·3941	6·30	·9147	7·30	·9962	8·30	·9999
4·35	·0085	5·35	·4306	6·35	·9253	7·35	·9968	8·35	·9999
4·40	·0121	5·40	·4671	6·40	·9347	7·40	·9973	8·40	·9999
4·45	·0167	5·45	·5034	6·45	·9431	7·45	·9977	8·45	1·0000
4·50	0·0226	5·50	0·5390	6·50	0·9506	7·50	0·9981		
4·55	·0301	5·55	·5739	6·55	·9572	7·55	·9984		
4·60	·0393	5·60	·6076	6·60	·9629	7·60	·9987		
4·65	·0505	5·65	·6401	6·65	·9680	7·65	·9989		
4·70	·0638	5·70	·6711	6·70	·9725	7·70	·9991		

On the basis of these evaluations we believe that the following figures should not be in error by more than one or two units in the last place:

$$\text{Mean } w = 5{\cdot}492\,0853, \quad \sigma_w = 0{\cdot}565\,992, \quad \mu_2 = 0{\cdot}320\,347,$$
$$\mu_3 = 0{\cdot}090\,935, \quad \mu_4 = 0{\cdot}353\,16, \quad \beta_1 = 0{\cdot}251\,53, \quad \beta_2 = 3{\cdot}4414.$$

From Table 1 we have obtained by backward interpolation the percentage points for w shown in Table 3. It was found that if the tables of standardized 0·5, 1·0, 2·5 and 5 % points of Pearson curves (Pearson & Hartley, 1954, Table 42) are used with the moments given above, we obtain almost complete agreement* to two decimal places with the figures given in Table 3. This means that to the accuracy considered, the distribution of w must be very closely representable by a Type VI or inverted beta curve.

Table 3. *Percentage points of range in samples of* $n = 200$

Percentage	0·1	0·5	1·0	2·5	5·0	10·0
Lower points	4·094	4·281	4·372	4·518	4·648	4·806
Upper points	7·670	7·225	7·020	6·731	6·496	6·238

Finally it seemed worth examining whether the values of the percentage points for w/s given by David *et al.* (1954, p. 491) would be altered substantially now that more accurate values of the moments of w were available. Using the same method as these authors (p. 483) we derived the values shown for comparison in Table 4. For practical purposes, the discrepancies are not serious. We have not attempted to use our results to improve the empirical formulae suggested by Tukey (1955). To get the best out of his method of attack further accurate computations would seem to be needed, perhaps as he suggests for $n = 1000$. To carry these out, the desk computing methods we have used, as described in the next section, would seem to involve a prohibitive amount of labour.

Table 4. *Percentage points of* w/s

	Lower percentage points				Upper percentage points			
	0·5	1·0	2·5	5·0	5·0	2·5	1·0	0·5
David *et al.*	4·50	4·56	4·67	4·78	6·38	6·59	6·85	7·03
From improved moments	4·53	4·59	4·68	4·78	6·39	6·60	6·84	7·01

3. Method of computation

The probability integral of range in a normal sample of size n, $P(W) = \Pr\{w \leqslant W\}$, may be expressed in the form

$$P(W) = \left\{\int_{-\frac{1}{2}W}^{\frac{1}{2}W} z(x)\,dx\right\}^n + 2n\int_{\frac{1}{2}W}^{\infty} z(u)\left\{\int_{u-W}^{u} z(x)\,dx\right\}^{n-1} du, \tag{2}$$

where $z(x) = (2\pi)^{-\frac{1}{2}} e^{-\frac{1}{2}x^2}$. When $n = 200$ the main work of computation consists in evaluating the second integral in (2) or

$$400\int_{\frac{1}{2}W}^{\infty} z(u)\left\{\int_{u-W}^{u} z(x)\,dx\right\}^{199} du \tag{3}$$

by quadrature. As seven decimal accuracy in $P(W)$ was aimed at, it was necessary to calculate $z(u)\left\{\int_{u-W}^{u} z(x)\,dx\right\}^{199}$ for a series of values of u from $\frac{1}{2}W$ up to the point where the expression became less than 1 unit in the 10th decimal place. The argument interval for u used in the quadrature had sometimes to be as small as 0·025. Values of $z(u)$ and $\int_{u-W}^{u} z(x)\,dx$ to 15 decimal places were taken from the National Bureau of Standards (1953) tables of the normal probability function, the integral being

* Only one of the eight values differed by as much as 0·005.

raised to the 199th power using tables of 15 figure logarithms. The first term in equation (2) was obtainable without quadrature, using the same tables.

REFERENCES

David, H. A., Hartley, H. O. & Pearson, E. S. (1954). *Biometrika*, **41**, 482–93.
National Bureau of Standards (1953). *Tables of Normal Probability Functions*. Applied Mathematics Series, no. 23. Washington: U.S. Department of Commerce.
Pearson, E. S. (1932). *Biometrika*, **24**, 404–17.
Pearson, E. S. & Hartley, H. O. (1942). *Biometrika*, **32**, 301–10.
Pearson, E. S. & Hartley, H. O. (1954). *Biometrika Tables for Statisticians*, **1**. Cambridge University Press.
Ruben, H. (1954). *Biometrika*, **41**, 200–27.
Tippett, L. H. C. (1925). *Biometrika*, **17**, 364–87.
Tukey, J. W. (1955). *Biometrika*, **42**, 480–5.

Studies in the history of probability and statistics. V. A note on playing cards

By M. G. KENDALL
Research Techniques Unit, London School of Economics

1. In an earlier article in this series (Kendall, 1956) I referred briefly to the introduction of playing cards into Europe. Subsequent correspondence arising out of that reference suggests that it may be useful to expand a little what was there said about the impact of cards on gambling.

2. Playing cards as we know them to-day in western Europe can be traced back in a clear line of descent to the beginning of the fifteenth century; but from that point backwards their history becomes more and more vague and their genealogy more and more fabulous. Where they originated is unknown. Claims have been put forward on behalf of origins in China, India, Arabia and Egypt. It is at least equally possible that they were independently invented in Europe. From the first the pictorial representations on the cards were thoroughly Western and do not suggest, to my eye at least, any trace of Eastern descent such, for example, as does the rook in chess.* Gambling with paper tickets is said to have been known in China in the twelfth century, and it is possible that the idea of playing cards drifted across to Europe along one of the early trade routes; but for the translation of the idea into practice one does not need to look further afield than fourteenth-century Italy.

3. No mention of cards has been traced in the West before A.D. 1350, and the absence of reference in authors like Chaucer and Dante, who mentioned everything, shows that they cannot have been known much before that date. They seem to have spread in the Mediterranean countries and Germany fairly rapidly. Numerals are said to have been added to the picture cards at Venice in A.D. 1377. Cards are mentioned at Nürnberg in A.D. 1380. In A.D. 1397 there appeared an ordinance in Paris prohibiting play at various games, including cards, to that part of the population who were engaged in manufacturing. (The story that cards were invented to amuse the mad Charles VI of France is false, although packs were made for him in A.D. 1393.) Specimens of such early dates have not survived. By A.D. 1423, however, the card pack appears to have evolved into its modern form. San Bernardino, whose sermon served me in good stead in the previous article (1956), refers to *charticellae in quibus variae figurae pingantur*, and goes on to mention the four suits, the Kings, Queens, Valets and Chevaliers and, as I interpret him, the trump cards of the tarot pack.†

* Modern tarot packs are worthless evidence in this connexion. Under the influence of occultists, notably Court de Gebelin, who suggested in the eighteenth century that the trump cards incorporated the lost book of Thoth, and Eliphas Lévi, whose work on magic popularized the idea in the nineteenth century, tarot cards have acquired symbols such as sphinxes which are absent from earlier western cards.

† The early suits were clubs, coins (diamonds), cups (hearts) and swords (spades), which San Bernardino identifies with brutality, avarice, drunkenness and hatred. The court cards denote those who are outstanding in these vices. 'Presbyterorum et presbyterarum in numerabilem multitudinem esse volo: unde admittantur pusilli et magni, et feminae et masculi, periti et ignari, sapientes et stulti.' A hundred and fifty years later one John Northbrooke was inspired by this passage to coin the famous description of cards as 'the devil's picture books'.

Miss Gertrude Moakley (1956) has recently suggested that the trump cards were representations of the *Trionfi* described by Petrarch in one of his most popular poems.

4. The game of tarock (French *tarot*, Italian *tarocchi*) is still played in southern Europe in various versions and is probably the oldest card game. In its modern form the tarot pack consists of 56 ordinary cards divided into four suits, 21 trump cards and a wild card or joker, 78 cards in all. Fifteenth-century packs of this type exist. There are some early packs of even larger size, notably the Minchiate pack of 97 cards, 56 ordinary cards, 35 trump cards and 6 wild cards. Historians of playing cards also mention a very rare set of engravings known as the Mantegna tarot; but they are probably not by Mantegna, and I doubt very much whether they are tarot cards. They are engraved on sizeable but thin sheets of paper, and can hardly have been used in any sort of game involving shuffling and dealing; they are divided into five sets of ten, the first, for example, enumerating various social grades from the beggar to the Pope, the second giving the nine Muses and Apollo, the third giving ten branches of learning, and so on. My own opinion is that these cards were a teaching device and that the unknown inventor of the tarot pack copied some of them. Where he got the others is a mystery unless Miss Moakley is right in identifying them with Petrarch's *Trionfi*.

5. At some unknown point of time the tarot pack was simplified, or so I believe. The trump cards were dropped, and of the 56 ordinary cards one of the court cards was also dropped. (In most countries it was the Chevalier who was dropped, but in Spain it was the Queen, for reasons which it is interesting but unprofitable to speculate upon.) Thus there evolved the basic pack of 52 cards which is in general use to-day. The tarot pack survived independently, but in northern Europe is now mainly used for fortune-telling.

6. The student of the history of probability is interested in these matters only in two respects: the degree to which cards extended and encouraged gambling, and the reasons for the choice of the number of cards in the early tarot packs. From what San Bernardino says it seems that gambling began at a very early stage; presumably, as soon as a game came into existence, the adversaries began to wager on the outcome. However, extensive gambling with cards was of very slow development. Cardano mentions the game of primero, but early writers on chance confine themselves mainly to dice. The reasons, I think, were twofold: first, the permutational arithmetic required to deal with probabilities at cards was too complicated; secondly, cards were very expensive and dice much more common. Cards did not oust dice until the eighteenth century. A third reason, possibly, is that cards and backgammon involved more skill and had a higher social status. James I (1603) puts it rather well:

'As for sitting, or human pastimes—since they may at times supply the room which, being empty, would be patent to pernicious idleness—I will not therefore agree with the curiosity of some learned men of our age in forbidding cards, dice and such like games of hazard; when it is foul and stormy weather, then I say, may ye lawfully play at the cards or tables; for, as to dicing, I think it becometh best deboshed soldiers to play at on the heads of their drums, being only ruled by hazard, and subject to knavish cogging; and as for the chess, I think it over-fond because it is overwise and philosophic a folly.'

James, apparently, was not very good at chess, but his balanced broadmindedness in a Puritan age compels respect.

7. It is interesting to inquire why the early tarot packs consisted of $56+21+1$ or $56+35+6$ cards, but before dabbling in numerology let us note how treacherous a subject it is. One can make out a very good case for a connexion between the modern pack of playing cards and the calendar. The four suits correspond to the four seasons, the thirteen cards in a suit correspond to the lunar months, the 52 cards to the weeks of the year. If we score 11 for a Knave, 12 for a Queen, and 13 for a King, and add all the points for the 52 cards we get 364, which, adding one for the joker, gives us the number of days in the year.* Many a historical point has been argued on less coincidental evidence than this, but there is, in fact, no connexion between the playing pack and the calendar. The early suit cards were 56 in number.

8. Nevertheless, there are some striking resemblances between the number of the cards in the early packs and the number of ways in which dice can fall. The main number, 56, is the number of ways in which three dice can be thrown, permutations excluded, and I pointed out in my previous article (1956) that these ways were well known by the fourteenth century. The number 21, likewise, is the number of ways in which two dice can be thrown, permutations excluded. The number 35 of the larger Minchiate tarot could arise either as the number of ways of throwing three dice when sixes are ignored, or the number of ways of throwing four (four-sided) astragali. I am not prepared to lean very heavily on these coincidences; but they suggest that perhaps the constructors of the first packs, having to choose somehow, were influenced by their knowledge of dice-throwing.

* Nor is this the worst. The number of letters in the sequence Ace, Two, Three, ..., Ten, Jack, Queen, King, is 52; so also in the sequence As, Deux, Trois, ..., Dix, Valet, Reine, Roi and in As, Zwei, Drei, ... Zehn, Bube, Dame, Konig, the ch being taken as one letter. I owe this information to the firm of Thos. de la Rue and Co. Ltd.

9. In conclusion, I should like to correct one guess made in my earlier article. I suggested that the game of hazard was brought back to Europe by the third crusaders. It may, indeed, have been brought back in some such way, but if so, must have been imported by earlier crusaders. The word *hasart* occurs in line 10557 of Wace's *Le Roman De Brut*, dated A.D. 1155, and also in Chrétien de Troyes' *Erec et Enide*, line 356, dated A.D. 1160–70. For these references I am indebted to Prof. Brian Woledge, who remarks, incidentally, that the appearance of the initial 'h' in 'hazard' is an etymological mystery which has never been solved.

REFERENCES

JAMES I (1603). *Basilikon Doron, or a King's Christian Duty towards God.*

KENDALL, M. G. (1956). Studies in the history of probability and statistics. II. The beginnings of a calculus of probabilities. *Biometrika*, **43**, 1.

MOAKLEY, GERTRUDE (1956). The tarot trumps and Petrarch's *Trionfi*. *Bull. N.Y. Publ. Lib.* **60**, 55.

A singularity in the estimation of binomial variance

BY ALAN STUART

Research Techniques Unit, London School of Economics

SUMMARY. For the symmetrical binomial distribution, the limit distribution of the sample variance is non-normal and has variance of order $1/n^2$.

1. For the binomial distribution, the sample proportion of 'successes', p, is a sufficient estimator of the probability parameter π, and

$$y = p(1-p)\,n/(n-1) \tag{1}$$

is an unbiased estimator of $\pi(1-\pi)$, which is n times the sampling variance of p, where n is sample size. $p(1-p)$ is the sample variance. Since y is a function of the sufficient statistic, it is the minimum variance unbiased estimator of its expectation. Its sampling variance is

$$V(y) = \left(\frac{n}{n-1}\right)^2 \{V(p) + V(p^2) - 2C(p, p^2)\},$$

which is expressible in terms of the moments about the origin of p as

$$V(y) = \left(\frac{n}{n-1}\right)^2 \{\mu_4' - 2\mu_3' + \mu_2' - (\mu_2' - \mu_1')^2\}.$$

On substitution for these moments we find

$$V(y) = \frac{\pi(1-\pi)}{n}\left\{(1-2\pi)^2 + \frac{2\pi(1-\pi)}{n-1}\right\}. \tag{2}$$

If $\pi \neq \frac{1}{2}$, (2) gives

$$V(y \mid \pi \neq \tfrac{1}{2}) \sim \pi(1-\pi)\,(1-2\pi)^2/n, \tag{3}$$

while if $\pi = \frac{1}{2}$ (2) becomes

$$V(y \mid \pi = \tfrac{1}{2}) = 1/\{8n(n-1)\}. \tag{4}$$

It is easily confirmed that the right-hand sides of (3) and (4) are the information bounds to the sampling variance of an estimator of $\pi(1-\pi)$, to order n^{-1}, n^{-2} respectively in the two cases.

The fact that at $\pi = \frac{1}{2}$ the variance of y is of lower order than n^{-1} raises the question whether its limiting distribution is normal in that case. In the following sections it is shown that this is not so.

2. The characteristic function of $p(1-p)$ may be written

$$\begin{aligned}\phi_{p(1-p)}(t) &= E\{\exp[\theta(p-p^2)]\} \\ &= \exp(\theta\pi^2)\,E\{\exp[\theta p(1-2\pi)]\exp[-\theta(p-\pi)^2]\},\end{aligned} \tag{5}$$

where $\theta = it$. Since, for random variables u, v,

$$E(uv) = E(u)\,E(v) + C(u, v),$$

we may replace the expectation of the two exponential terms in (5) by the product of their expectations plus their covariance

$$C = C\{\exp[\theta p(1-2\pi)],\quad \exp[-\theta(p-\pi)^2]\}. \tag{6}$$

If $\pi = \frac{1}{2}$, the first variate on the right of (6) is a constant, so

$$C = 0 \quad (\pi = \tfrac{1}{2}). \tag{7}$$

If $\pi \neq \frac{1}{2}$, we may use differentials in (6) to obtain

$$C = \{\theta(1-2\pi)\exp[\theta p(1-2\pi)]\}\{-\theta\exp[-\theta(p-\pi)^2]\}C\{p,(p-\pi)^2\} + R_n, \tag{8}$$

the remainder being of lower order in n than the leading term given on the right of (8). The first two factors on the right of (8) are to be evaluated at the true parameter point $p = \pi$, and are there of order zero in n. The third factor is

$$C\{p,(p-\pi)^2\} = E(p-\pi)^3 = \pi(1-\pi)(1-2\pi)/n^2$$

exactly, so we may write (8) as

$$C = kn^{-2} + o(n^{-2}). \tag{9}$$

Using (7) and (9), (5) becomes, for all π,

$$\phi_{p(1-p)}(t) = \exp(\theta\pi^2)\,\phi_{p(1-2\pi)}(t)\,\phi_{-(p-\pi)^2}(t) + o(n^{-\frac{3}{2}}). \tag{10}$$

3. By the classical central limit theorem,

$$(p-\pi)/\{\pi(1-\pi)/n\}^{\frac{1}{2}}$$

has a limiting normal distribution with characteristic function $\exp(\frac{1}{2}\theta^2)$. It follows that its square has a limiting χ^2 distribution with 1 degree of freedom and characteristic function $(1-2\theta)^{-\frac{1}{2}}$. From these results, it follows that

$$\left.\begin{aligned}\phi_{p(1-2\pi)}(t) &= \exp\{\theta\pi(1-2\pi) + \tfrac{1}{2}\theta^2\pi(1-\pi)(1-2\pi)^2/n\}(1+o(1)),\\ \phi_{-(p-\pi)^2}(t) &= \{1+2\theta\pi(1-\pi)/n\}^{-\frac{1}{2}}(1+o(1)).\end{aligned}\right\} \tag{11}$$

Substitution of (11) into (10) gives

$$\phi_{p(1-p)}(t) = \exp\{\theta\pi(1-\pi) + \tfrac{1}{2}\theta^2\pi(1-\pi)(1-2\pi)^2/n\}\{1+2\theta\pi(1-\pi)/n\}^{-\frac{1}{2}}(1+o(1)). \tag{12}$$

From the definition of y at (1), we have, correcting for the mean

$$\phi_{y-\pi(1-\pi)}(t) = \exp\{\theta\pi(1-\pi)/(n-1) + \tfrac{1}{2}\theta^2\pi(1-\pi)(1-2\pi)^2\,n/(n-1)^2\} \times\{1+2\theta\pi(1-\pi)/(n-1)\}^{-\frac{1}{2}}(1+o(1)). \tag{13}$$

If $\pi \neq \frac{1}{2}$, we standardize y, using (3), to obtain for

$$z = \{y-\pi(1-\pi)\}/\{V(y)\}^{\frac{1}{2}}$$

the characteristic function

$$\phi_z(t) = \exp\{\theta[n\pi(1-\pi)]^{\frac{1}{2}}/[(n-1)(1-2\pi)] + \tfrac{1}{2}\theta^2n^2/(n-1)^2\} \times\{1+2\theta[n\pi(1-\pi)]^{\frac{1}{2}}/[(n-1)(1-2\pi)]\}^{-\frac{1}{2}}(1+o(1)), \tag{14}$$

so that

$$\lim_{n\to\infty}\phi_z(t) = \exp(\tfrac{1}{2}\theta^2) \quad (\pi \neq \tfrac{1}{2}), \tag{15}$$

and the distribution of y, properly standardized, is normal in the limit.

If, in (13), $\pi = \frac{1}{2}$, the term in θ^2 disappears and (13) becomes

$$\phi_{y-\frac{1}{4}}(t) = \exp\{\theta/[4(n-1)]\}\{1+\theta/[2(n-1)]\}^{-\frac{1}{2}}(1+o(1)). \tag{16}$$

We now standardize by (4) and obtain from (16)

$$\phi_z(t) = \exp\{\theta(n/[2(n-1)])^{\frac{1}{2}}\}\{1+\theta[2n/(n-1)]^{\frac{1}{2}}\}^{-\frac{1}{2}}(1+o(1)). \tag{17}$$

Thus

$$\lim_{n\to\infty}\phi_z(t) = \exp(\theta/2^{\frac{1}{2}})(1+\theta.2^{\frac{1}{2}})^{-\frac{1}{2}} \quad (\pi = \tfrac{1}{2}), \tag{18}$$

and the distribution is non-normal.

From the form of (18), in fact, it is clear that

$$1 - z.2^{\frac{1}{2}}$$

has a χ^2 distribution with 1 degree of freedom in large samples.

4. The reason for this singularity in the distribution of y is clear. When $\pi = \frac{1}{2}$, $\pi(1-\pi)$ attains its maximum value $\frac{1}{4}$. y is an unbiased estimator of this value, but itself has a maximum value $\frac{1}{4}n/(n-1)$, which approaches $\frac{1}{4}$ when n becomes large. Since y is thus estimating a value which tends to one extreme of its range, it is to be expected that the distribution of y should be very skewed, and the effect becomes more pronounced as sample size increases.

The position is comparable with that of the distribution of the squared sample multiple correlation coefficient, R^2, in samples from a multinormal population. If the population multiple correlation parameter is non-zero, R^2 is asymptotically normal with variance of order n^{-1}. But if the parameter is zero, R^2, like y in our discussion above, is essentially estimating a parameter at one extreme of its range, has variance of order n^{-2}, and is non-normally distributed.

5. Finally, an implication of the result should be briefly mentioned. In testing the equality of the values of π in two populations, we use a large sample standard error test based on the fact that

$$(p_1-p_2)\Big/\left\{\pi(1-\pi)\left(\frac{1}{n_1}+\frac{1}{n_2}\right)\right\}^{\frac{1}{2}}$$

is asymptotically a standardized normal variate. If we are testing the composite hypothesis, π being unspecified, we estimate $\pi(1-\pi)$ by y as above, calculated from the pooled samples, with $n = n_1+n_2$. Our result implies that, if $\pi = \frac{1}{2}$, y estimates its expectation with greater precision than if $\pi \neq \frac{1}{2}$, and this presumably improves the accuracy of the normal approximation using the estimated standard error.

Student's distribution and Riemann's elliptic geometry

By AUREL WINTNER
Johns Hopkins University

The density of probability for Student's ratio is*

$$c/(1+x^2)^{\frac{1}{2}k}, \tag{1}$$

where $k\ (>1)$ is the number of the variables and the value of $c = c_k$ is determined by the condition that the integral of the function (1) over the x-line is 1. The algebraic structure of the function (1) suggests a simple geometrical approach† to Student's distribution. In fact, such an approach leads to the following interpretation:

If k is denoted by $n+2$, and if R_n (where $n>0$, hence $k>2$) is the space of Riemann's n-dimensional elliptic geometry, then refer R_n to his normal co-ordinates $x_1, \ldots, x_n$, where $-\infty<x_i<\infty$; consider on R_n that distribution which represents equidistribution (in terms of the co-ordinates x_i); denote by L a line through the origin of R_n; finally, denote by x, where $-\infty<x<\infty$, the abscissa of that point of L which is the orthogonal projection of an arbitrary point $(x_1, \ldots, x_n)$ of R_n. *Then the function* (1) *on* L *is the density of probability of the orthogonal projection of the equidistribution on* R_n. This can be seen as follows:

Stereographic projection of the Euclidean n-space shows that the squared line-element on R_n is

$$ds^2 = g\sum_{i=1}^{n} dx_i^2, \quad \text{where} \quad g^{\frac{1}{2}} = 1+\sum_{i=1}^{n} x_i^2, \tag{2}$$

if the 'diameter' of R_n is chosen to be the unit of length. But if

$$ds^2 = \sum_{i=1}^{n}\sum_{j=1}^{n} g_{ij}(x_1, \ldots, x_n)\,dx_i dx_j$$

is any (positive-definite) Riemannian metric, then the Riemannian volume element is $(\det g_{ik})^{\frac{1}{2}}$ times the Euclidean volume element. Hence it is clear from (2), where $g_{ij} = g\delta_{ij}$, that the Riemannian volume of that portion (infinite slab of thickness dx) of the $(x_1, \ldots, x_n)$-space which is contained between the two hyperplanes $x_n = x$, $x_n = x+dx$ is dx times the value of the $(n-1)$-fold integral

$$\int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty}(1+x^2+r^2)^{-n}\,dx_1\ldots dx_{n-1}, \tag{3}$$

where $r^2 = x_1^2+\ldots+x_{n-1}^2$. Accordingly, the function (3) of x, where $-\infty<x<\infty$, is the density of the equidistribution projected on L, if L is chosen to be the x_n-axis.

In order to evaluate the integral, use (ordinary) polar co-ordinates, r and $n-2$ angles, in the Euclidean $(n-1)$-space $(x_1, \ldots, x_{n-1})$. These reduce the $(n-1)$-fold integral (3) to a constant multiple of an integral over r alone. In fact, the result is

$$C\int_0^{\infty}(1+x^2+r^2)^{-n}r^{n-2}\,dr,$$

* See, for example, Kendall (1952, chapter 10) where further references will be found.

† In this connexion see Wintner (1940, pp. 287–97 and 1947, pp. 168–73).

where $C = C_{n-2}$, being the contribution of the $n-2$ angles, is the Euclidean measure of the unit sphere ($r = 1$). But if x, hence $X = (1+x^2)^{\frac{1}{2}}$, is fixed and if the integration variable r is replaced by t, where $r = Xt^{\frac{1}{2}}$, then the last integral appears in the form

$$\int_0^\infty X^{-2n}(1+t)^{-n}(\tfrac{1}{2}X^{n-1})\,t^{\frac{1}{2}(n-3}\,dt,$$

which, since X is independent of t, is a constant multiple of $X^{-2n}X^{n-1} = 1/X^{n+1}$ or, since $X = (1+x^2)^{\frac{1}{2}}$, of Student's density (1), where $k = n+2$.

REFERENCES

KENDALL, M. G. (1952). *The Advanced Theory of Statistics*, **1**, 5th ed. London: Charles Griffin and Co. Ltd.

WINTNER, A. (1940). *Astronomical Papers Dedicated to Elis Strömgren*. Copenhagen.

WINTNER, A. (1947). *The Fourier Transforms of Probability Distributions*. Baltimore.

Some interrelations among compound and generalized distributions

BY JOHN GURLAND

Iowa State College

1. INTRODUCTION

It is well known (cf. Greenwood & Yule, 1920; Feller, 1943) that the Pascal (negative binomial distribution) may be regarded as a compound distribution if the compounding is effected through a gamma distribution. More recently it has been noted (cf. Jones & Mollison, 1948; Quenouille, 1949) that the Pascal distribution may also be regarded as a generalized Poisson distribution if the generalizing is effected through a logarithmic distribution.

The development in the present paper consists of the following: first, a convenient symbolism formalizing the notions of compound and generalized distributions is introduced; then a simple relation that exists for a certain class of compound and generalized distributions is formulated in a theorem. Some examples relating to the theorem are noted. Next, the relation involving the Poisson, gamma and logarithmic distributions indicated in the papers referred to above, is extended one stage further, and the general extension for a finite number of stages is indicated.

2. COMPOUND AND GENERALIZED DISTRIBUTIONS

Compound and generalized Poisson variables have been studied by Feller (1943) and Satterthwaite (1942). More recently, compound and generalized variables other than Poisson have also been considered (Skellam, 1952; Feller, 1950; Gurland, 1957). The present definitions and notation are given to facilitate manipulations with such distributions.

Definition 1. *Compound distribution*

Let the random variable X_1 have the distribution function $F_1(x_1 \mid \theta)$ for a given value of the parameter θ. Suppose now that θ is regarded as a random variable X_2, say, with distribution function $F_2(x_2)$. Denote by $X_1 \wedge X_2$ the random variable with distribution function

$$\int_{-\infty}^{\infty} F_1(x_1 \mid cx_2)\,dF_2(x_2), \tag{1}$$

where c is a constant which is arbitrary or restricted in some prescribed sense. Then the random variable $X_1 \wedge X_2$ (uniquely defined here except for the constant c) is called a compound X_1 variable with respect to the 'compounder' X_2.

The advantage of introducing the constant c will become apparent in the formulation of the theorem. It should also be remarked that in the above definition the constant could be incorporated in the F_2 term instead of in the F_1 term, since (1) can be written equivalently as

$$\int_{-\infty}^{\infty} F_1(x_1 \mid x_2)\,dF_2\left(\frac{x_2}{c}\right).$$

If several parameters $\theta_1, \theta_2, \ldots, \theta_r$, say, appear in the distribution function F_1, then the notation $X_1 \wedge X_2$ is inadequate to specify which of the parameters is involved in the compounding operation. In such a case, to avoid ambiguity, we would write

$$X_1(\theta_1, \theta_2, \ldots, \theta_r) \underset{\theta_1}{\wedge} X_2,$$

where θ_1, say, is the parameter specifically involved in the compounding operation.

As an example of definition 1, let X_1 be a Poisson variable with p.g.f. (probability generating function) $e^{\theta(z-1)}$, and X_2 be a gamma variable with c.f. (characteristic function) $(1-it/\alpha)^{-\lambda}$. Then the p.g.f. of $X_1 \wedge X_2$ is given by

$$\frac{\alpha^\lambda}{\Gamma(\lambda)} \int_0^\infty e^{cx(z-1)} e^{-\alpha x} x^{\lambda-1} dx = \left[1 - \frac{c}{\alpha}(z-1)\right]^{-\lambda} \quad (c>0,\ \alpha>0,\ \lambda>0). \tag{2}$$

If we consider a Pascal variable with p.g.f.

$$[1-p(z-1)]^{-k} \quad (p>0,\ k>0), \tag{3}$$

it is evident that for each set of values c, α, λ in (2) a set of values p, k can be assigned in (3) such that expressions (2) and (3) have the same value whatever be z. Also, to any set of values p, k in (3) there correspond values c, α, λ (not unique here) such that (2) and (3) have the same value whatever be z. We shall call two such distributions equivalent and now give the following formal definition.

Definition 2. *Equivalent distributions*

Suppose the random variables X_1, X_2 have distribution functions $F_1(x \mid \alpha)$, $F_2(x \mid \beta)$ respectively. (α and/or β may be multidimensional.) If for each α there exists some β and for each β there exists some α such that $F_1(x \mid \alpha) = F_2(x \mid \beta)$ whatever be x, the random variables X_1 and X_2 are said to be equivalent, and we write $X_1 \sim X_2$.

Occasionally it is convenient to represent a random variable by the name of its corresponding distribution. Thus, in the case of Poisson, gamma and Pascal distributions the meaning of the relation

$$\text{Poisson} \wedge \text{gamma} \sim \text{Pascal} \tag{4}$$

should be clear.

The following definition of a generalized distribution is given formally in terms of probability generating functions because of greater generality and ease of manipulation.

Definition 3. *Generalized distribution*

Let the random variables X_1, X_2 have p.g.f.'s $g_1(z)$, $g_2(z)$ respectively. Denote by $X_1 \vee X_2$ the random variable with p.g.f. $g_1(g_2(z))$. Then $X_1 \vee X_2$ is called a generalized X_1 variable with respect to the 'generalizer' X_2.

It may be remarked here that X_1, X_2 need not be discrete variables as assumed, for instance, by Feller (1943). For X_1 the p.g.f. is merely taken as Ez^{X_1}, and likewise, of course, for X_2.

The following theorem gives a simple relation for a certain class of compound and generalized distributions.

THEOREM. *Let X_1 be a random variable with p.g.f. $[h(z)]^\theta$, where θ is a given parameter. Suppose now θ is regarded as a random variable X_2, say, with distribution function F_2 and p.g.f. g_2. Then, whatever be X_2*

$$X_1 \wedge X_2 \sim X_2 \vee X_1.$$

Proof. The p.g.f. of $X_1 \wedge X_2$ is given by

$$\int_{-\infty}^{\infty} [h(z)]^{cx}\, dF_2(x)$$

and that of $X_2 \vee X_1$ is given by

$$g_2(g_1(z)) = \int_{-\infty}^{\infty} [h(z)]^{\theta x}\, dF_2(x).$$

It is obvious, of course, that these p.g.f.'s exist, at least, for $z = e^{it}$, and are equal when $c = \theta$; hence the theorem is proved.

Some examples illustrating the above theorem are the following:

$$\text{Poisson} \wedge \text{Poisson} \sim \text{Poisson} \vee \text{Poisson}, \tag{5}$$

$$\underset{k}{\text{Pascal}}\,(k,p) \wedge \text{Poisson} \sim \text{Poisson} \vee \text{Pascal}, \tag{6}$$

$$\underset{k}{\text{Pascal}}\,(k,p) \wedge \text{gamma} \sim \text{gamma} \vee \text{Pascal}. \tag{7}$$

It is, of course, apparent from the theorem that two different mathematical interpretations may underlie the same distribution. Thus, as already pointed out by Feller (1943) the Neyman Type A contagious distribution may be interpreted as a compound Poisson or a generalized Poisson distribution. This is epitomized in (5) above.

Relation (7) is also interesting in that the gamma variable appearing in the right-hand member is a continuous random variable.

3. Extension of a certain relation

As mentioned in § 1 it is known that the following equivalence holds:

$$\text{Poisson} \wedge \text{gamma} \sim \text{Poisson} \vee \text{logarithmic}. \tag{8}$$

An immediate extension of (8) is the following:

$$\underset{k}{\text{Pascal}}\,(k,p) \wedge \text{gamma} \sim \text{Pascal} \vee \text{logarithmic}. \tag{9}$$

To see this, let X_i $(i = 1, 2)$ be Pascal variables with p.g.f.'s $[1-p_i(z_i-1)]^{-k_i}$ $(i = 1, 2)$; let X_3 be a logarithmic variable with p.g.f. $-\alpha \log(1-\tau z)$ and X_4 be a gamma variable with c.f. $(1-it/\beta)^{-\lambda}$. The p.g.f. $g(z)$, say, of $\underset{k_1}{X_1(p_1, k_1)} \wedge X_4$ is given by

$$\begin{aligned} g(z) &= \frac{\beta^\lambda}{\Gamma(\lambda)} \int_0^\infty [1-p_1(z-1)]^{-cx} e^{-\beta x} x^{\lambda-1} dx \\ &= \left[1 + \frac{c}{\beta} \log(1+p_1) + \frac{c}{\beta} \log\left(1 - \frac{p_1}{1+p_1} z\right)\right]^{-\lambda}. \end{aligned}$$

On the other hand, the p.g.f. $h(z)$, say, of $X_2 \vee X_3$ is given by

$$h(z) = [1 + p_2 + \alpha p_2 \log(1-\tau z)]^{-k_2},$$

and the required equivalence $X_1 \wedge X_4 \sim X_2 \vee X_3$ is immediately apparent.

It is also possible to extend (9) in the same manner as (8) by letting the gamma and logarithmic variables play the same role as in the previous extension. This extension could be carried through any finite number n, say, of stages. One might refer to the new random variables obtained in this manner as nth stage compound Poisson and nth stage generalized Poisson respectively. In such a context the Pascal distribution could be referred to as a first stage compound Poisson or generalized Poisson distribution; whereas the variable $\underset{k}{\text{Pascal}}\,(k, p) \wedge \text{gamma}$ or $\text{Pascal} \vee \text{logarithmic}$ could be referred to as second-stage variables of this type.

To sketch the method of proof for the general case let Y_i be an ith stage compound Poisson and W_i be an ith stage generalized Poisson variable. Let the corresponding compounder and generalizer be gamma (β_i, λ_i) and logarithmic (τ_i) respectively.

Repeated application of the compounding operation in (2) shows that the p.g.f.

$$\begin{aligned} &\text{of } Y_1 \text{ is } \quad \left[1 - \frac{c_1}{\beta_1}(z-1)\right]^{-\lambda_1}, \\ &\text{of } Y_2 \text{ is } \quad \left[1 + \frac{c_2}{\beta_2} \log\left\{1 - \frac{c_1}{\beta_1}(z-1)\right\}\right]^{-\lambda_2}, \\ &\text{of } Y_3 \text{ is } \quad \left[1 + \frac{c_3}{\beta_3} \log\left\{1 + \frac{c_2}{\beta_2} \log\left(1 - \frac{c_1}{\beta_1}(z-1)\right)\right\}\right]^{-\lambda_3}, \\ &\text{etc.} \end{aligned}$$

To write the p.g.f. of W_i let the original Poisson variable have p.g.f. $e^{\theta(z-1)}$ and write the p.g.f. of logarithmic (τ_i) as $-\alpha_i \log(1-\tau_i z)$, where $\alpha_i \log(1-\tau_i) = -1$ $(0 < \tau_i < 1)$.

Then the p.g.f.

$$\begin{aligned} &\text{of } W_1 \text{ is } \quad e^{-\theta}(1-\tau_1 z)^{-\alpha_1\theta}, \\ &\text{of } W_2 \text{ is } \quad e^{-\theta}[1 + \tau_1\alpha_2 \log(1-\tau_2 z)]^{-\alpha_1\theta}, \\ &\text{of } W_3 \text{ is } \quad e^{-\theta}[1 + \tau_1\alpha_2 \log\{1 + \tau_2\alpha_3 \log(1-\tau_3 z)\}]^{-\alpha_1\theta}, \\ &\text{etc.} \end{aligned}$$

Since

$$\begin{aligned} e^{-\theta} &= (1-\tau_1)^{\alpha_1\theta} \quad (0 < \tau_1 < 1) \\ &= (1+\pi_1)^{-\alpha_1\theta}, \text{ say} \quad (\pi_1 > 0), \end{aligned}$$

it is clear that for each finite $k = 1, 2, \ldots$, the p.g.f.'s of Y_k and W_k are of the same form and hence yield equivalent random variables.

4. SUMMARY

A simple relation between a certain class of compound and generalized distributions is pointed out and a few examples are given. In particular, some distributions may be regarded both as compound and as generalized. Finally, a relation between the Poisson, gamma and logarithmic distributions is extended, which involves the Pascal distribution, and a generalization of this extension is indicated.

REFERENCES

FELLER, W. (1943). On a general class of contagious distributions. *Ann. Math. Statist.* **14**, 389–400.

FELLER, W. (1950). *An Introduction to Probability Theory and its Applications.* New York: John Wiley and Sons.

GREENWOOD, M. & YULE, G. UDNY (1920). An inquiry into the nature of frequency distribution representative of multiple happenings with particular reference to the occurrence of multiple attacks of disease or repeated accidents. *J. R. Statist. Soc.* **83**, 255–79.

GURLAND, J. (1957). On Beall and Rescia's generalization of Neyman's contagious distributions. Submitted for publication.

JONES, P. C. T. & MOLLISON, J. E. (1948). A technique for the quantitative estimation of soil micro-organisms. *J. Gen. Micr.* **2**, 54–69.

QUENOUILLE, M. H. (1949). A relation between the logarithmic, Poisson, and negative binomial series. *Biometrics*, **5**, 162–4.

SATTERTHWAITE, F. E. (1942). Generalized Poisson distribution. *Ann. Math. Statist.* **13**, 410–17.

SKELLAM, J. G. (1952). Studies in statistical ecology. I. Spatial pattern. *Biometrika*, **39**, 346–62.

A note on tests of significance for linear functional relationships.

BY M. S. BARTLETT

University of Manchester

Williams (1955) has recently developed further the exact significance tests for a single hypothetical non-null canonical variate, first proposed by him (1952) in special cases and considered further by the author (1951). He has also, however, advocated analogous tests for hypothetical linear relations or null variates in situations where these tests are not correct, and the purpose of the present note is to draw attention to this.

It seems convenient to summarize these tests in terms of the factorizations discussed in my 1951 paper. For simplicity the case of one null variate will be considered, in contrast with one non-null variate (the more general case being r null variates, in contrast with r non-null variates). It was pointed out in my earlier paper that if the regression relation of a vector variate $\mathbf{x}$ with p components x_i $(i = 1, \ldots, p)$ on another variate $\mathbf{y}$ with q components $(p \leqslant q)$ gives the p canonical roots R_j^2 $(j = 1, \ldots, p)$, so that

$$\Lambda(n, p, q) = \prod_j (1 - R_j^2), \tag{1}$$

and if a hypothetical linear combination ξ of the x_i gives the corresponding measure of association (more strictly, disassociation) $1 - \phi$, then the exact criterion

$$\Lambda'(n-1, p-1, q) = \Lambda/(1-\phi)$$

may be factorized further into

$$\Lambda' = \frac{(1-R_1^2)}{1-\phi}\,[\prod_{j>1} (1-R_j^2)], \tag{2}$$

where the two factors in (2) may be tested *approximately*. In place of the 'approximate' factorization in (2), 'exact' factorizations are possible

$$\left.\begin{aligned} \Lambda' &= \Lambda''(n-1, p-1, 1)\,\Lambda'''(n-2, p-1, q-1), \\ \Lambda' &= \Lambda^{\mathrm{IV}}(n-1, p-1, q-1)\,\Lambda^{\mathrm{V}}(n-q, p-1, 1), \end{aligned}\right\} \tag{3}$$

in terms of the projection of ξ in the sample space of the q variates y_k $(k = 1, \ldots, q)$.

Consider now the case of one hypothetical null variate η (in relation to the set $\mathbf{y}$), and let the corresponding measure of association of this with the other set $\mathbf{y}$ be $1 - \lambda$. Williams suggests that η will define

$p-1$ variates orthogonal to η in the sample space, **z**, say, and that if the relation of **z** with **y** is measured by $\Lambda'(n, p-1, q)$ we may consider

$$\Lambda''(n-p+1, 1, q) = \Lambda(n, p, q)/\Lambda'(n, p-1, q). \tag{4}$$

If, moreover, the relation with **x** (in the reduced space orthogonal to **z**) of the projection ζ, say, of **z** on **y** is given by $\Lambda'''(n-p+1, 1, p-1)$, then the further factorization of Λ'' in (4) is possible,

$$\Lambda'' = \Lambda'''(n-p+1, 1, p-1)\,\Lambda^{\text{iv}}(n-2p+2, 1, q-p+1), \tag{5}$$

with an alternative factorization (the remaining variables in **y** being taken first)

$$\Lambda'' = \Lambda^{\text{v}}(n-p+1, 1, q-p+1)\,\Lambda^{\text{vi}}(n-q, 1, p-1). \tag{6}$$

One possible minor point of obscurity may be cleared. It is not immediately obvious that the total number of degrees of freedom in the criterion $\Lambda'(n, p-1, q)$ in (4) is still n, as the variates in **z** are already orthogonal to η. However, as by hypothesis η is randomly orientated with respect to **y**, the *stochastic* part of **z** will be randomly orientated with respect to **y**, and there is no reduction in n. Nevertheless, this perhaps indicates the sort of difficulty that will arise, for in general, of course, **z** is not random with respect to **y**, and under such non-null conditions for **z** the relations of the variates are presumably more complicated.

That this is so may be demonstrated in a simple case, but first it is instructive to consider the more direct factorization corresponding to the factorization in (2) for testing ξ. The simple factor $1-\lambda$ is of the type $\Lambda(n, 1, q)$ and is expected to be insignificant. Moreover, we may write

$$1-\lambda = \left[\frac{1-\lambda}{1-R_p^2}\right](1-R_p^2), \tag{7}$$

where the significance of R_p^2 would indicate the existence of *no* null variate, and the significance of the other factor that η is not the right null-variate. Unfortunately, while the overall test based on $1-\lambda$, when available, is exact (on the usual normality assumptions), the two separate factors in (7) are rather less convertible into approximate χ^2 components than in (2), for the crudest χ^2 testing of R_p^2 would rely on *all* the remaining $p-1$ variates **z** having close relation with **y**, and this need not be the case. However, as actual insignificance of some of the other canonical correlations would be evidence of more than one null variate (cf. the discussion by Tintner (1946) and Bartlett (1948) of two null variates), a cautious use of such χ^2 approximations may still be convenient if a further factorization of $1-\lambda$ as in (7) is being considered.

The difficulty with any exact analogues of the factorizations in (3) is that, while we may eliminate $p-1$ hypothetical canonical variates if these are given *a priori* (and so test the goodness of fit of these variates), the choice of $p-1$ variates orthogonal to η *in the sample space*, when only η is given *a priori*, is not equivalent to this.

It might be noticed that if in (7) the condition $p \leqslant q$ were dropped and the case $q = 1$, say, considered, R_p is zero, and no factorization of $1-\lambda$ arises. In contrast, Williams gives an analysis of variance table (top of p. 375) in which the sum of squares λ for η is tested not against a residual sum of squares $1-\lambda$, but against the residual $1-R^2$, a quantity quite irrelevant to the analysis of η and one, moreover, which could be made as near zero as we wished merely by increasing the *irrelevant* association with **y** of the $p-1$ other canonical variates. A very special case which may bring this point out even more clearly is that of $p = 2$, when there are two dependent variates x_1 and x_2, of which x_1, say, is the hypothetical null variate. While it might be of interest to know whether x_2 is also uncorrelated with **y**, there is no justification for inflating the straightforward significance of x_1, as measured by λ, if x_2 is correlated with **y**.

I learn from correspondence with Dr Williams that he fully agrees with my criticisms, and he notes that §§ VII and VIII of his paper will consequently require amendment.

REFERENCES

Bartlett, M. S. (1948). A note on the statistical estimation of supply and demand relations from time-series. *Econometrica*, **16**, 323.

Bartlett, M. S. (1951). The goodness of fit of a single hypothetical discriminant function in the case of several groups. *Ann. Eugen., Lond.*, **16**, 198.

Tintner, G. (1946). Multiple regression for systems of equations. *Econometrica*, **14**, 5.

Williams, E. J. (1952). Some exact tests in multivariate analysis. *Biometrika*, **39**, 17.

Williams, E. J. (1955). Significance tests for discriminant functions and linear functional relationships. *Biometrika*, **42**, 360.

The moments of the Leipnik distribution

By M. G. KENDALL

Research Techniques Unit, London School of Economics

1. The distribution

$$dF = \frac{(1-r^2)^{\frac{1}{2}n-\frac{1}{2}}\,dr}{B(\frac{1}{2}n+\frac{1}{2},\frac{1}{2})\,(1+\rho^2-2\rho r)^{\frac{1}{2}n}} \quad (-1 \leqslant r \leqslant 1) \tag{1}$$

was obtained by Leipnik (1947), following a method due to Madow (1945), as an approximation to the distribution of the first serial correlation, circularly defined, with known mean, in samples of n from a Markoff scheme. Recent work by Daniels (1956) and Jenkins (1956) show that approximations to the distributions of the coefficient with fitted mean, and non-circularly defined, have moments which are simply expressible in terms of the moments of (1).

Leipnik himself found the mean and second moment of the distribution by a complicated procedure. Jenkins (1956) has recently obtained two more by a method which is also rather complicated. In this note I give a general expression which enables moments of all orders to be written down.

2. We have

$$\begin{aligned}
-\frac{1}{n}\frac{\partial \mu'_k}{\partial \rho} &= \int_{-1}^{1} \frac{r^k(\rho-r)\,(1-r^2)^{\frac{1}{2}n-\frac{1}{2}}\,dr}{B(\frac{1}{2}n+\frac{1}{2},\frac{1}{2})\,(1+\rho^2-2\rho r)^{\frac{1}{2}n+1}} \\
&= \int \frac{\rho r r^{k-1}}{1+\rho^2-2\rho r}\,dF - \int \frac{1+\rho^2-(1+\rho^2-2\rho r)}{2\rho(1+\rho^2-2\rho r)}\,r^k\,dF \\
&= \frac{1}{2}\int \frac{1+\rho^2-(1+\rho^2-2\rho r)}{1+\rho^2-2\rho r}\,r^{k-1}\,dF - \frac{1+\rho^2}{2\rho}\int \frac{r^k\,dF}{1+\rho^2-2\rho r} + \frac{1}{2\rho}\mu'_k \\
&= \frac{1}{2\rho}\mu'_k - \tfrac{1}{2}\mu'_{k-1} + \frac{1+\rho^2}{2\rho}\int \frac{(\rho-r)\,r^{k-1}}{1+\rho^2-2\rho r}\,dF,
\end{aligned}$$

whence

$$\left(\frac{1}{n}\frac{\partial}{\partial \rho}+\frac{1}{2\rho}\right)\mu'_k = \left(\frac{1+\rho^2}{2\rho n}\frac{\partial}{\partial \rho}+\frac{1}{2}\right)\mu'_{k-1}. \tag{2}$$

3. This relation can be used to build up the moments from $\mu_0\,(=1)$ and $\mu'_1\,(=n\rho/(n+2))$. But we can also use it to derive explicit expressions. We note in the first place that if P_k is a polynomial in ρ of degree k the solution of

$$\left(\frac{1}{n}\frac{\partial}{\partial \rho}+\frac{1}{2\rho}\right)X = P_k$$

is a polynomial of order $k+1$. It follows that μ'_k is a polynomial in ρ of order k. Moreover, it follows from (2) and the lower values of μ'_k that even-order moments contain only even powers of ρ and odd-order moments only odd powers of ρ.

Let

$$\mu'_k = \sum_{m=0}^{k} a_{km}\rho^m.$$

On differentiating (2) m times and putting $\rho = 0$ we find

$$a_{km} = \frac{1}{n+2m}\{(m+1)\,a_{k-1,m+1} + (n+m-1)\,a_{k-1,m-1}\}. \tag{3}$$

4. We have $\mu_0 = 1$, $a_{00} = 1$, and hence

$$a_{1,1} = \frac{1}{n+2}n,$$

giving

$$\mu'_1 = \frac{n\rho}{n+2}. \tag{4}$$

By successive application of (3) we find

$$\mu'_2 = \frac{1}{n+2} + \frac{n(n+1)\rho^2}{(n+2)(n+4)}, \tag{5}$$

$$\mu'_3 = \frac{3n\rho}{(n+2)(n+4)} + \frac{n(n+1)\rho^3}{(n+4)(n+6)}, \tag{6}$$

$$\mu'_4 = \frac{3}{(n+2)(n+4)} + \frac{6n(n+1)\rho^2}{(n+2)(n+4)(n+6)} + \frac{n(n+1)(n+3)\rho^4}{(n+4)(n+6)(n+8)}, \tag{7}$$

$$\mu'_5 = \frac{15n\rho}{(n+2)(n+4)(n+6)} + \frac{10n(n+1)\rho^3}{(n+4)(n+6)(n+8)} + \frac{n(n+1)(n+3)\rho^5}{(n+6)(n+8)(n+10)}, \tag{8}$$

$$\mu'_6 = \frac{15}{(n+2)(n+4)(n+6)} + \frac{45n(n+1)\rho^2}{(n+2)(n+4)(n+6)(n+8)} + \frac{15n(n+1)(n+3)\rho^4}{(n+4)(n+6)(n+8)(n+10)} + \frac{n(n+1)(n+3)(n+5)\rho^6}{(n+6)(n+8)(n+10)(n+12)}. \tag{9}$$

Formulae (6) and (7) agree with Jenkins's results. The general law of formation of the terms will now be clear. The coefficients may be set out as follows:

Order of moment	Coefficient of powers										
	0	1	2	3	4	5	6	7	8	9	10
1	.	1	.	.	.	.	.	.	.	.	.
2	1	.	1	.	.	.	.	.	.	.	.
3	.	3	.	1	.	.	.	.	.	.	.
4	3	.	6	.	1	.	.	.	.	.	.
5	.	15	.	10	.	1	.	.	.	.	.
6	15	.	45	.	15	.	1	.	.	.	.
7	.	105	.	105	.	21	.	1	.	.	.
8	105	.	420	.	210	.	28	.	1	.	.
9	.	945	.	1260	.	378	.	36	.	1	.
10	945	.	4725	.	3150	.	630	.	45	.	1

The numerical coefficient of a_{km} is $\dfrac{k!}{m!\{\frac{1}{2}(k-m)\}!\,2^{\frac{1}{2}(k-m)}}$. The results are easily demonstrated by induction from (3).

5. From (2) we find that the moment generating function M' obeys the relation

$$\left(\frac{1}{n}\frac{\partial}{\partial\rho} + \frac{1}{2\rho}\right)\frac{\partial M'}{\partial\theta} = \left(\frac{1+\rho^2}{2\rho n}\frac{\partial}{\partial\rho} + \frac{1}{2}\right)M'.$$

Putting

$$M(\theta)' = \exp\left(\frac{n\rho\theta}{n+2}\right)M(\theta),$$

we obtain the m.g.f. of the mean moments. Identifying coefficients in θ^k we find

$$\left(\frac{\partial}{\partial\rho} + \frac{n}{2\rho}\right)\mu_{k+1} = \frac{n+2-(n-2)\rho^2}{2\rho(n+2)}\frac{\partial\mu_k}{\partial\rho} - \frac{nk}{n+2}\mu_k + \frac{nk}{n+2}\left\{\frac{n+2-(n-2)\rho^2}{2\rho(n+2)}\right\}\mu_{k-1}. \tag{10}$$

Since $\mu_0 = 1$, $\mu_1 = 0$ we find

$$\mu_2 = \frac{1}{n+2} - \frac{n(n-2)\rho^2}{(n+2)^2(n+4)} \sim \frac{1-\rho^2}{n}, \tag{11}$$

$$\mu_3 = -\frac{6n\rho}{(n+2)^2(n+4)} + \frac{2n(n-2)(3n-2)\rho^3}{(n+2)^3(n+4)(n+6)} \sim \frac{-6\rho(1-\rho^2)}{n^2}, \tag{12}$$

$$\mu_4 = \frac{3}{(n+2)(n+4)} - \frac{6n(n^2-8n+4)\rho^2}{(n+2)^3(n+4)(n+6)} + \frac{3n(n-2)(n^3-14n^2+12n-8)\rho^4}{(n+2)^4(n+4)(n+6)(n+8)} \sim \frac{3(1-\rho^2)^2}{n^2}. \tag{13}$$

These also agree with Jenkins's results. It does not appear that general formulae are obtainable with the simplicity of the moments about zero.

6. The limit of (10) with n large may be written

$$n\mu_{k+1} = (1-\rho^2)\frac{\partial \mu_k}{\partial \rho} - 2\rho k\mu_k + k(1-\rho^2)\mu_{k-1}. \tag{14}$$

Thus, with $k = 4, 5$, we have

$$\mu_5 \sim \frac{-60\rho(1-\rho^2)^2}{n^3}, \tag{15}$$

$$\mu_6 \sim \frac{15(1-\rho^2)^3}{n^3}. \tag{16}$$

Again the general law is clear. In standard measure the odd-order moments tend to zero, the even-order moments to those of the normal distribution; and thus the distribution tends to normality.

7. It may be of interest to record another recurrence relation between the moments, namely,

$$\mu'_{k+1}(n) = \mu'_{k-1}(n) + \frac{n+1}{n+2}\{2\rho\mu'_k(n+2) - (1+\rho^2)\mu'_{k-1}(n+2)\}, \tag{17}$$

where $\mu'(n)$ refers to the moment about the origin based on a sample of n.

REFERENCES

DANIELS, H. E. (1956). The approximate distribution of serial correlation coefficients. *Biometrika*, **43**, 169.

JENKINS, G. M. (1956). Tests of hypotheses in the linear autoregressive model. II. *Biometrika*, **43**, 186.

LEIPNIK, R. B. (1947). Distribution of the serial correlation coefficient in a circularly correlated universe. *Ann. Math. Statist.* **18**, 86.

MADOW, W. G. (1945). Note on the distribution of the serial correlation coefficient. *Ann. Math. Statist.* **16**, 308.

The effect of transformations of variables upon their correlation coefficients

BY M. H. QUENOUILLE

Research Techniques Unit, London School of Economics

Suppose x_t to be normally distributed with zero mean and standard deviation σ, and let y_t be derived from x_t by a monotonic transformation

$$y_t = \sum_{i=1}^{\infty} a_i H_i\left(\frac{x_t}{\sigma}\right),$$

where $H_i(x)$ is the ith Hermite polynomial. Let the correlation between x_t and x_{t-s} be ρ_s.

The correlation between y_t and y_{t-s} may then be derived using the relations

$$E\left[H_i\left(\frac{x_t}{\sigma}\right)H_j\left(\frac{x_{t-s}}{\sigma}\right)\right] = 0 \quad (i \neq j)$$
$$= i!\rho_s^i \quad (i=j).$$

These little-known relationships spring from the more-familar equations

$$E\left[H_i\left(\frac{x_t}{\sigma}\right)H_j\left(\frac{x_t}{\sigma}\right)\right] = 0 \quad (i \neq j)$$
$$= i! \quad (i=j),$$

by transforming the variables to independent normal functions, x_t, $\rho_s x_i + y_t$, and using

$$H_j(y_t + c) = \sum_{k=0}^{j}\binom{j}{k} c^k H_{j-k}(y_t).$$

It follows that

$$\operatorname{cov}(y_t, y_{t-s}) = \sum_{i=1}^{\infty} a_i^2 i!\rho_s^i$$

and

$$\rho(y_t, y_{t-s}) = \frac{a_1^2\rho_s + 2a_2^2\rho_s^2 + 6a_3^2\rho_s^3 + \ldots}{a_1^2 + 2a_2^2 + 6a_3^2 + \ldots}.$$

Consequently, $\rho(y_t, y_{t-s})$ is a weighted average of ρ_s, ρ_s^2, ρ_s^3, ... and, except when $\rho_s = 1$, is necessarily less than ρ_s. Further, the difference between $\rho(y_t, y_{t-s})$ and ρ_s will be greater for negative ρ_s than for positive ρ_s.

It follows that, in any time series, transformation of the observations to normality maximizes the serial correlations, and that failure to achieve normality may be marked by the positive serial correlations being larger in modulus than the negative serial correlations.

This latter fact will be most obvious when the serial correlations contain a harmonic component of amplitude b. In this instance, the average of ρ_s over a run of values of s is zero for odd n, and

$$b(n-1)(n-3)\ldots/\{n(n-2)\ldots\}$$

for even n. Thus the average of $\rho(y_t, y_{t-s})$ over a run of values of s is

$$b\frac{a_2^2+(1.3)^2a_4^2+(1.3.5)^2a_6^2+\ldots}{a_1^2+2a_2^2+6a_3^2+24a_4^2+\ldots}.$$

This expression may be fairly appreciable. For instance, for the transformation $y_t = \exp(x_t/\sigma) - \exp 0{\cdot}5$, $a_i = \sigma^{2i}/i!$, and this expression becomes

$$b\frac{\left(\frac{\sigma^2}{2}\right)^2+\frac{1}{(2!)^2}\left(\frac{\sigma^2}{2}\right)^4+\frac{1}{(3!)^2}\left(\frac{\sigma^2}{2}\right)^6+\ldots}{\sigma^2+\frac{1}{2!}\sigma^4+\frac{1}{3!}\sigma^6+\ldots}.$$

The numerator of this expression may be derived from $I_0(\sigma^2)$, or, more rapidly in most instances, by direct calculation. For example, for $\sigma = 1$, this expression is equal to $0{\cdot}155b$, an appreciable shift in the mean value.

Further properties of an angular transformation of the correlation coefficient

By B. I. HARLEY

University College, London

1. In a previous paper (1956) the properties of the transformation

$$y = \sin^{-1} r,$$

where r is the correlation coefficient in a sample of size n from a normal bivariate population with correlation ρ, were discussed and it was shown that if $\mathscr{E}_n(\sin^{-1} r)$ denotes the expectation of $\sin^{-1} r$ then

$$\mathscr{E}_n(\sin^{-1} r) = \sin^{-1}\rho \tag{1}$$

for all values of n. This was proved directly only for even values of n, but by considering the form of the distribution of $r/(1-r^2)^{\frac{1}{2}}$ we can obtain directly an alternative method to that given in the previous paper, of proving (1) for odd values of n. The results suggest that the distribution of $r/(1-r^2)^{\frac{1}{2}}$ may be similar to that of the non-central t distribution, and this I have investigated in another paper (Harley, 1957).

2. (x_i, y_i) for $i = 1, 2, \ldots, n$, are n pairs of values drawn randomly from a normal bivariate population with correlation ρ. Without loss of generality we can arrange that in the population the mean values of x and y are zero, and the standard deviations are unity. If $\bar{x}$ and $\bar{y}$ are the means of the n sample values, then the coefficient of correlation of the sample is given by

$$r = \frac{\sum_{i=1}^{n}(x_i-\bar{x})(y_i-\bar{y})}{\left(\sum_{i=1}^{n}(x_i-\bar{x})^2\sum_{i=1}^{n}(y_i-\bar{y})^2\right)^{\frac{1}{2}}}. \tag{2}$$

Transforming to the variables

$$X = \frac{x-\rho y}{(1-\rho^2)^{\frac{1}{2}}} \quad \text{and} \quad Y = y,$$

we find that

$$X^2 + Y^2 = (x^2 - 2\rho xy + y^2)/(1-\rho^2),$$

and (X_i, Y_i) may be considered as n pairs of values drawn randomly from a normal bivariate population with zero means, unit standard deviations and correlation zero. Writing $\bar{X}$, $\bar{Y}$ for the sample means of X, Y and S_X^2, S_Y^2 for their variances, it follows from (2) that

$$\begin{aligned}
\frac{r}{(1-r^2)^{\frac{1}{2}}} &= \frac{\sum_{i=1}^{n}(x_i-\bar{x})(y_i-\bar{y})}{\left\{\sum_{i=1}^{n}(x_i-\bar{x})^2\sum_{i=1}^{n}(y_i-\bar{y})^2-\left(\sum_{i=1}^{n}(x_i-\bar{x})(y_i-\bar{y})\right)^2\right\}^{\frac{1}{2}}} \\
&= \frac{(1-\rho^2)^{\frac{1}{2}}\sum_{i=1}^{n}(X_i-\bar{X})(Y_i-\bar{Y})+\rho\sum_{i=1}^{n}(Y_i-\bar{Y})^2}{(1-\rho^2)^{\frac{1}{2}}\left\{\sum_{i=1}^{n}(X_i-\bar{X})^2\sum_{i=1}^{n}(Y_i-\bar{Y})^2-\left[\sum_{i=1}^{n}(X_i-\bar{X})(Y_i-\bar{Y})\right]^2\right\}^{\frac{1}{2}}} \\
&= \frac{\sum_{i=1}^{n}(X_i-\bar{X})(Y_i-\bar{Y})}{\left\{\sum_{i=1}^{n}(X_i-\bar{X})^2\sum_{i=1}^{n}(Y_i-\bar{Y})^2-\left[\sum_{i=1}^{n}(X_i-\bar{X})(Y_i-\bar{Y})\right]^2\right\}^{\frac{1}{2}}} \\
&\qquad + \frac{\rho}{(1-\rho^2)^{\frac{1}{2}}}\frac{\left(\sum_{i=1}^{n}(Y_i-\bar{Y})^2\right)^{\frac{1}{2}}}{\left(\sum_{i=1}^{n}(X_i-\bar{X})^2\right)^{\frac{1}{2}}}\left(1-\frac{\left[\sum_{i=1}^{n}(X_i-\bar{X})(Y_i-\bar{Y})\right]^2}{\sum_{i=1}^{n}(X_i-\bar{X})^2\sum_{i=1}^{n}(Y_i-\bar{Y})^2}\right)^{-\frac{1}{2}} \\
&= \frac{r_0}{(1-r_0^2)^{\frac{1}{2}}}+\frac{\rho}{(1-\rho^2)^{\frac{1}{2}}}\left(\frac{S_Y^2}{S_X^2}\right)^{\frac{1}{2}}\frac{1}{(1-r_0^2)^{\frac{1}{2}}},
\end{aligned}$$

where r_0 is the correlation coefficient of a sample drawn randomly from a normal bivariate population with zero correlation.

Let $$\frac{r_0}{(1-r_0^2)^{\frac{1}{2}}} = t_0 \quad \text{and} \quad \frac{S_Y^2}{S_X^2} = F,$$

then $$\frac{1}{(1-r_0^2)^{\frac{1}{2}}} = (1+t_0^2)^{\frac{1}{2}}$$

and $$\frac{r}{(1-r^2)^{\frac{1}{2}}} = t_0 + F^{\frac{1}{2}}(1+t_0^2)^{\frac{1}{2}}\tan(\sin^{-1}\rho).$$

Let $$\sin^{-1}\rho = \theta \quad \text{and} \quad F^{\frac{1}{2}}\tan\theta = z,$$

then $$\frac{r}{(1-r^2)^{\frac{1}{2}}} = t_0 + z(1+t_0^2)^{\frac{1}{2}}. \tag{3}$$

3. If $$E_n = \mathscr{E}_n(\sin^{-1} r),$$

then $$\begin{aligned}
E_n &= \mathscr{E}_n\left(\tan^{-1}\frac{r}{(1-r^2)^{\frac{1}{2}}}\right) \\
&= \mathscr{E}_n(\tan^{-1}[t_0+z(1+t_0^2)^{\frac{1}{2}}]), \\
\frac{\partial E_n}{\partial\theta} &= \mathscr{E}_n\frac{F^{\frac{1}{2}}\sec^2\theta}{\{(1+z^2)(1+t_0^2)^{\frac{1}{2}}+2t_0 z\}}.
\end{aligned}$$

F and t_0 are independent for the case we are considering, and thus to derive the expectation we can integrate first with respect to t_0 and then with respect to F.

The distribution of t_0 is known to be

$$p(t_0) = \frac{1}{B(\frac{1}{2}n-1, \frac{1}{2})(1+t_0^2)^{\frac{1}{2}n-\frac{1}{2}}}.$$

We wish to show that $E_3 = \sin^{-1}\rho = \theta$. If $n = 3$

$$p(t_0) = \frac{1}{\pi(1+t_0^2)}, \quad p(F) = \frac{1}{(1+F)^2}.$$

Hence $$\frac{\partial E_3}{\partial\theta} = \int_0^\infty \frac{F^{\frac{1}{2}}\sec^2\theta}{\pi}\left[\int_{-\infty}^{+\infty}\frac{dt_0}{\{(1+z^2)(1+t_0^2)^{\frac{1}{2}}+2t_0 z\}(1+t_0^2)}\right]p(F)\,dF.$$

Let $t_0 = \sinh\phi$; then the integral with respect to t_0 becomes

$$\begin{aligned} I &= \int_{-\infty}^{+\infty} \frac{\cosh\phi\, d\phi}{\{(1+z^2)\cosh\phi + 2z\sinh\phi\}\cosh^2\phi} \\ &= \int_{-\infty}^{+\infty} \frac{\operatorname{sech}^2\phi\, d\phi}{1+z^2+2z\tanh\phi} \\ &= \left[\frac{1}{2z}\log[(1+z^2)+2z\tanh\phi]\right]_{-\infty}^{+\infty} \\ &= \frac{1}{2z}\log\frac{(1+F^{\frac{1}{2}}\tan\theta)^2}{(1-F^{\frac{1}{2}}\tan\theta)^2} \\ &= \frac{2}{z}\tanh^{-1}(F^{\frac{1}{2}}\tan\theta). \end{aligned}$$

Hence
$$\frac{\partial E_3}{\partial\theta} = \int_0^{\infty} \frac{2\sec^2\theta}{\pi\tan\theta}\tanh^{-1}(F^{\frac{1}{2}}\tan\theta)(1+F)^{-2}\,dF.$$

Integrating by parts we find that $\partial E_3/\partial\theta = 1$. Thus $E_3 = \sin^{-1}\rho + C$, where C is a constant. When $\rho = 0$ we have

$$\begin{aligned} E_3 &= \int_{-1}^{+1} \frac{\sin^{-1}r}{\pi}(1-r^2)^{-\frac{1}{2}}\,dr \\ &= \frac{1}{\pi}[(\sin^{-1}r)^2]_{-1}^{+1} - \frac{1}{\pi}\int_{-1}^{+1}\sin^{-1}r(1-r^2)^{-\frac{1}{2}}\,dr = 0. \end{aligned}$$

Thus $C = 0$ and, in general, $E_3 = \sin^{-1}\rho$.

In a previous paper (1956) we deduced the following recurrence relationship:

$$\frac{1}{(1-\rho^2)^{\frac{1}{2}}} - \frac{\partial}{\partial\rho}\mathscr{E}_{n-1}(\sin^{-1}r) = \frac{\rho}{(1-\rho^2)}(n-2)\,[\mathscr{E}_{n-1}(\sin^{-1}r) - \mathscr{E}_{n+1}(\sin^{-1}r)]. \tag{4}$$

If $n-1 = 3$, we have just seen that

$$\mathscr{E}_{n-1}(\sin^{-1}r) = \mathscr{E}_3(\sin^{-1}r) = \sin^{-1}\rho.$$

Consequently the left-hand side of equation (4) is zero and it follows that

$$\mathscr{E}_3(\sin^{-1}r) = \mathscr{E}_5(\sin^{-1}r) = \sin^{-1}\rho.$$

Substituting $(n-1) = 5, 7, 9, \ldots$ in (4) in turn, we see that

$$\sin^{-1}\rho = \mathscr{E}_3(\sin^{-1}r) = \mathscr{E}_5(\sin^{-1}r) = \mathscr{E}_7(\sin^{-1}r) = \ldots.$$

We have shown previously that

$$\sin^{-1}\rho = \mathscr{E}_4(\sin^{-1}r) = \mathscr{E}_6(\sin^{-1}r) = \mathscr{E}_8(\sin^{-1}r) = \ldots.$$

Hence
$$\mathscr{E}_n(\sin^{-1}r) = \sin^{-1}\rho \quad \text{for} \quad n \geqslant 3.$$

REFERENCE

HARLEY, B. I. (1956). *Biometrika*, **43**, 219.
HARLEY, B. I. (1957). *Biometrika*, **44**, 219.

Heterogeneity of error variances in a randomized block design

BY FRANKLIN A. GRAYBILL AND JOHN LEROY FOLKS

Oklahoma Agricultural and Mechanical College

1. INTRODUCTION

In a randomized block experiment we frequently wish to test the hypothesis that all the treatment means are equal. When we have heterogeneity of error variances, however, the ratio of the treatment mean square to the error mean square is not distributed as Snedecor's F. For example, consider a randomized

block design with p treatments occurring on each of r blocks. If each of the first n_1 treatments have variance σ_1^2 and each of the next n_2 treatments have variance σ_2^2, etc., and if

$$\sum_{i=1}^{g} n_i = p,$$

the mathematical model may be written as

$$Y_{ijk} = \mu + t_{ij} + b_k + e_{ijk} \quad (i = 1, 2, \ldots, g;\ j = 1, 2, \ldots, n_i;\ k = 1, 2, \ldots, r), \tag{1·1}$$

where the e_{ijk}'s are assumed to be normally distributed such that

$$\mathscr{E}(e_{ijk}) = 0 \text{ for all } i, j, \text{ and } k; \quad \mathscr{E}(e_{ijk}^2) = \sigma_i^2 \text{ for all } j \text{ and } k$$

and

$$\mathscr{E}(e_{ijk}e_{smn}) = 0 \quad \text{unless} \quad i = s,\ j = m \text{ and } k = n.$$

Graybill (1954) has given an exact test, based on Hotelling's T^2, for testing the hypothesis,

$$H_0\colon\ t_{11} = t_{12} = \ldots = t_{1n_1} = t_{21} = \ldots = t_{gn_g}.$$

This test is valid if $r > p - 1$ and requires inverting a matrix of order $p - 1$.

The purpose of this paper is to give an exact test of H_0 which is simpler than that given by Graybill when $n_i > 1$ for at least one i. If $n_1 = n_2 = \ldots = n_g = 1$, then the test now proposed reduces to that given by Graybill. The latter is valid when $r > p - 1$ regardless of the values of n_i, but the new method is valid if $r > g - 1$ and requires the inversion of a matrix of order $g - 1$. Therefore, if $n_i > 1$ for at least one i, the method proposed in this paper is considerably simpler than the earlier method.

2. The test criterion

Using the ith subset of observations* let us conduct an analysis of variance as in Table 1 for each subset for which $n_i > 1$.

Table 1. *Analysis of variance for i-th subset*

Source of variation	Degrees of freedom	Sum of squares
Blocks	$r - 1$	$n_i \sum_k (y_{i.k} - y_{i..})^2 = A_i$
Treatments	$n_i - 1$	$r \sum_j (y_{ij.} - y_{i..})^2 = B_i$
Error	$(r-1)(n_i-1)$	$\sum_{jk} (Y_{ijk} - y_{i.k} - y_{ij.} + y_{i..})^2 = C_i$

The ratio $\dfrac{B_i}{n_i - 1} \Big/ \dfrac{C_i}{(r-1)(n_i-1)} = F_i$ (where $Y_{ij.}$ indicates Y_{ijk} summed over k and $y_{ij.}$ indicates the average of Y_{ijk} when summed over k, etc.) is distributed as Snedecor's F with $(n_i - 1)$ and $(r-1)(n_i-1)$ degrees of freedom if and only if the hypothesis H_i: $(t_{i1} = t_{i2} = \ldots = t_{in_i})$ is true.

If we let $q - 1$ represent the exact number of subsets of Y_{ijk} in which $n_i > 1$, then we will run an analysis of variance as in Table 1 for each of these subsets. Also since Y_{ijk} is independent of Y_{smn} if $s \neq i$ (regardless of the values of j, k, m and n), these $q - 1$ analyses of variance will yield $q - 1$ independent F_i each distributed as F with the appropriate degrees of freedom if H_i $(i = 1, 2, \ldots, q-1)$ is true. For definiteness let us assume that the first $q - 1$ subsets each have $n_i > 1$. That is to say, suppose $n_1 > 1, n_2 > 1, \ldots, n_{q-1} > 1$. This is no way affects the generality of the discussion and the result is that if the hypotheses

$$H_1\colon\ (t_{11} = t_{12} = \ldots = t_{1n_1});\quad H_2\colon\ (t_{21} = t_{22} = \ldots = t_{2n_2});\quad \ldots;\quad H_{q-1}\colon\ (t_{q-1,1} = t_{q-1,2} = \ldots = t_{q-1,n_{q-1}})$$

are each true, then $F_1, F_2, \ldots, F_{q-1}$ are independently distributed as F with degrees of freedom

$$(n_1 - 1),\ (r-1)(n_1-1);\quad (n_2-1),\ (r-1)(n_2-1);\quad \ldots;\quad (n_{q-1}-1),\ (r-1)(n_{q-1}-1)$$

respectively.

* By the ith subset of observations we will mean all observations where the first subscript is equal to i. For example the second subset of observations consists of Y_{2jk} where $j = 1, 2, \ldots, n_2$; $k = 1, 2, \ldots, r$.

If we average the model (1·1) over the subscript j we get

$$\sum_{j=1}^{n_i} \frac{Y_{ijk}}{n_i} = \mu + \sum_{j=1}^{n_i} \frac{t_{ij}}{n_i} + b_k + \sum_{j=1}^{n_i} \frac{e_{ijk}}{n_i}. \tag{2·1}$$

If we denote $\sum_{j=1}^{n_i} Y_{ijk}/n_i$ by B_{ik} and let $T_i = \sum_j t_{ij}/n_i$ and $d_{ik} = \sum_j e_{ijk}/n_i$, then we can write (2·1) as

$$B_{ik} = \mu + T_i + b_k + d_{ik}. \tag{2·2}$$

From the assumptions in (1·1)

$$\mathscr{E}(d_{ij}) = \mathscr{E}\sum_j e_{ijk}/n_i = 0, \quad \mathscr{E}(d_{ij}^2) = \sigma_i^2/n_i,$$

$$\mathscr{E}(d_{ij}d_{is}) = 0 \text{ for } j \neq s; \quad \mathscr{E}(d_{ij}d_{vj}) = 0 \text{ for } i \neq v.$$

Thus (2·2) is the model considered by Graybill (1954) and we can use Hotelling's T^2 to test the hypothesis $T_1 = T_2 = \ldots = T_g$. The procedure is to define the $g \times 1$ column vector X_j by

$$X_j = \begin{bmatrix} B_{1j} - B_{gj} \\ B_{2j} - B_{gj} \\ \vdots \\ B_{g-1,j} - B_{gj} \end{bmatrix}$$

and define the $g \times 1$ vector $\bar{X} = \sum_{j=1} X_j/r$. Then if H_0 is true, the quantity

$$F_0 = \frac{r(r-g+1)}{g-1} \bar{X}'[\sum_j (X_j - \bar{X})(X_j - \bar{X})']^{-1}\bar{X}$$

is distributed as F with $r-1$ and $r-g+1$ degrees of freedom (if $r > g-1$).

We will now show that $F_1, F_2, \ldots, F_q$ are jointly independently distributed as F if H_0 is true.

Let $(y_{ij.} - y_{i..}) = u_{ij}$ and $(y_{ijk} - y_{i.k} - y_{ij.} + y_{i..}) = v_{ijk}$. Since $F_1, F_2, \ldots, F_{q-1}$ are each functions of u_{ij} and v_{ijk}, and since F_q is a function of B_{mn}, and since u_{ij}, v_{ijk} and B_{mn} are all normally distributed, it follows that $F_1, F_2, \ldots, F_{q-1}$, F_q are jointly independent if $\operatorname{cov}(B_{mn}, u_{ij}) = \operatorname{cov}(B_{mn}, v_{ijk}) = 0$ for all i, j, k, m and n.

It is obvious that $\operatorname{cov}(B_{mn}, u_{ij}) = 0$ for $m \neq i$. Let us consider the case when $m = i$:

$$\begin{aligned} \operatorname{cov}(B_{in}, u_{ij}) &= \mathscr{E}(e_{i.n})(e_{ij.} - e_{i..}) \\ &= \mathscr{E}\frac{e_{ijn}^2}{rn_i} - \mathscr{E}\sum_j \frac{e_{ijn}^2}{rn_i^2} = \frac{\sigma_i^2}{rn_i} - \frac{\sigma_i^2}{rn_i} = 0. \end{aligned}$$

Also it is immediately clear that $\operatorname{cov}(B_{mn}, v_{ijk}) = 0$, for $m \neq i$. Let us consider the case when $m = i$:

$$\begin{aligned} \operatorname{cov}(B_{in}, v_{ijk}) &= \mathscr{E}(e_{i.n})(e_{ijk} - e_{ij.} - e_{i.k} + e_{i..}) \\ &= \mathscr{E}\frac{e_{ijk}^2}{n_i} - \mathscr{E}\frac{e_{ijk}}{rn_i} - \mathscr{E}\sum_j \frac{e_{ijk}^2}{n_i^2} + \mathscr{E}\sum_j \frac{e_{ijk}^2}{rn_i^2} \\ &= \frac{\sigma_i^2}{n_i} - \frac{\sigma_i^2}{rn_i} - \frac{\sigma_i^2}{n_i} + \frac{\sigma_i^2}{rn_i} = 0. \end{aligned}$$

We have, therefore, that the F_i ($i = 1, 2, \ldots, q$) are jointly independent and each is distributed as F if, and only if, H_0 is true; i.e. if, and only if, $t_{11} = t_{12} = \ldots = t_{1n_1} = t_{21} = \ldots = t_{gn_g}$. We can test H_0, therefore, by combining q independent tests of significance.

In a recent paper Birnbaum (1954) has summarized various methods of combining independent tests of significance including Fisher's test (Fisher, 1946, § 21·1). If Fisher's test is used, then the procedure for testing H_0 is to run the $q-1$ analyses of variance (for which $n_i > 1$) and tabulate the significance level of each F_i value; call these significance levels $P_1, P_2, \ldots, P_{q-1}$. If $r > g-1$ then F_q can be obtained and the significance level P_q tabulated (if $r \leqslant g-1$ the test proposed in this paper breaks down). If H_0 is true then

$$P = -\sum_{i=1}^{q} 2\log_e P_i$$

is distributed as a χ^2 variate with $2q$ degrees of freedom.

REFERENCES

Birnbaum, A. (1954). *J. Amer. Statist. Ass.* **49**, 559.

Fisher, R. A. (1946). *Statistical Methods for Research Workers.* Edinburgh: Oliver and Boyd.

Graybill, F. A. (1954). *Biometrics*, **10**, 516.

An extension property of a class of balanced incomplete block designs

By G. P. SILLITTO

Imperial Chemical Industries Ltd, Metals Division

Let a balanced incomplete block design which enables v varieties or treatments to be compared in b blocks (each block accommodating k varieties, each variety occurring once in each of r blocks, and each pair of varieties occurring together in λ of the blocks) be represented by a matrix having v rows and b columns, in which the element in the ith row and jth column is $+1$ if the ith variety is tested in the jth block, and -1 if this variety is not tested in the jth block. Then it is not difficult to show that the rows of the matrix are orthogonal to each other if, and only if,

$$b = 4(r-\lambda). \tag{1}$$

This note is concerned with balanced incomplete block designs for which (1) holds, and in particular with pointing out a 'law of composition' which enables a third design to be obtained by a simple process from any two such designs.

Using the known relations for any balanced incomplete block design

$$rv = bk, \tag{2}$$

$$\lambda = r(k-1)/(v-1), \tag{3}$$

a little algebra shows that if (1) is to be true, then necessarily

$$2k = v \pm \sqrt{v}, \tag{4}$$

so that (1) can only hold if v is the square of an integer. If (1) is true, the two values for k given by (4) are the values for a design and its complement. However, (4) serves to determine only the ratios b/r and λ/r, through (2) and (3) respectively. Hence what may be called the 'simple multiples' of a design (v, b, k, r, λ), i.e. the spectrum of designs $(v, nb, k, nr, n\lambda)$ in which n is an integer or any fraction which makes nb, nr and $n\lambda$ all integral, are covered by the same conditions as have been discussed. The value of this observation arises from the fact that when a design allowed by (2) and (3) does not exist, or is difficult to find, some of its simple multiples often exist and can easily be found.

Designs conforming to (1), (2) and (3) are of the form

$$v = u^2, \quad b = 2uN, \quad k = \tfrac{1}{2}u(u \pm 1), \quad r = (u \pm 1)N, \quad \lambda = \tfrac{1}{2}(u \pm 2)N, \tag{5}$$

where u is an integer > 1; N is an integer $\geqslant \frac{1}{2}u$ such that b, r and λ are integers; and designs with the parameters corresponding to $+$ signs represent the complements of the designs with parameters corresponding to $-$ signs. From any of these designs which exist and can be found, other designs can be obtained by using the law of composition given in the next paragraph.

From any pair of known designs for which (1) is true there can be found a third design, by the following process. Call the matrices of the two known designs M_1 and M_2; then the direct product $M_1 \times M_2$ of these two matrices is obtained by substituting for each element $+1$ of M_1 the matrix M_2, and for each element -1 of M_1 the matrix $-M_2$. Then the matrix $M_3 = M_1 \times M_2$ has $v_3 = v_1 v_2$ rows and $b_3 = b_1 b_2$ columns, and all its elements are $+1$ or -1. A little consideration shows that all the rows of M_3 contain the same number $r_3 = r_1 r_2 + (b_1 - r_1)(b_2 - r_2)$ of $+1$ elements, and all the columns contain the same number $k_3 = k_1 k_2 + (v_1 - k_1)(v_2 - k_2)$ of $+1$ elements. Further consideration of the different kinds of row which occur in M_3 shows that in any pair of rows, the number of columns in which both rows have $+1$ elements is $\lambda_3 = 6r_1 r_2 - 8r_1\lambda_2 - 8r_2\lambda_1 + 12\lambda_1\lambda_2$. Hence M_3 is the matrix of a balanced incomplete block design with parameters v_3, b_3, k_3, r_3, λ_3.

It is easily found also that M_3 conserves the property of row-orthogonality, and hence it in turn can be used as a factor in forming the direct product matrix representing yet another design; and so on. Clearly also in the direct product $M_1 \times M_2$, M_2 can be identical with, or the complement of, M_1, if desired. Therefore, if one or more of the designs in the series (5) are known, an infinite number of balanced incomplete block designs can be constructed.

The first members of the series (5) are

u	N	v	b	k	r	λ	Notes
2	1	4	4	3	3	2	Known, unreduced design
3	2	9	12	3	4	1	Known, see Fisher & Yates (1953)
4	2	16	16	6	6	2	Known, see below
5	4	25	40	10	16	6	Not known, but see below
6	3	36	36	15	15	6	Known, see below

A design with $u = 4$ and $N = 2$ is well known, being given by Fisher & Yates (1953). Such a design can also be obtained by composition of the design for $u = 2$, $N = 1$ with its own complement by the direct-product-matrix method.

No design corresponding to $u = 5$, $N = 4$ appears to be known, but the $\frac{2}{3}$ simple multiple of it, i.e. a design with $u = 5$, $N = 6$, $v = 25$, $b = 60$, $k = 10$, $r = 24$, $\lambda = 9$, is obtained (using the terminology of Bose, 1939) by considering the modul of residue classes (mod 5) and letting five varieties u_1, u_2, u_3, u_4, u_5 correspond to any element u of the modul. Then the design is obtained by adding the elements of the modul to the following twelve initial blocks, keeping the suffixes invariant:

$(0_1, 1_1, 2_1, 3_1, 3_2, 4_2, 2_3, 1_4, 0_5, 4_5)$ $(0_1, 1_1, 2_1, 3_1, 3_2, 2_3, 4_3, 1_4, 4_4, 0_5)$
$(0_1, 4_1, 0_2, 1_2, 2_2, 3_2, 3_3, 4_3, 2_4, 1_5)$ $(0_1, 0_2, 1_2, 2_2, 3_2, 3_3, 2_4, 4_4, 1_5, 4_5)$
$(1_1, 0_2, 4_2, 0_3, 1_3, 2_3, 3_3, 3_4, 4_4, 2_5)$ $(1_1, 4_1, 0_2, 0_3, 1_3, 2_3, 3_3, 3_4, 2_5, 4_5)$
$(2_1, 1_2, 0_3, 4_3, 0_4, 1_4, 2_4, 3_4, 3_5, 4_5)$ $(2_1, 4_1, 1_2, 4_2, 0_3, 0_4, 1_4, 2_4, 3_4, 3_5)$
$(3_1, 4_1, 2_2, 1_3, 0_4, 4_4, 0_5, 1_5, 2_5, 3_5)$ $(3_1, 2_2, 4_2, 1_3, 4_3, 0_4, 0_5, 1_5, 2_5, 3_5)$
$(0_1, 1_1, 0_2, 1_2, 0_3, 1_3, 0_4, 1_4, 0_5, 1_5)$ $(0_1, 2_1, 0_2, 2_2, 0_3, 2_3, 0_4, 2_4, 0_5, 2_5)$

No design corresponding to $u = 6$, $N = 3$ appears to have been published, but one can be obtained by considering the modul of residue classes (mod 3, 3) and letting four varieties u_1, u_2, u_3, u_4 correspond to every element u of the modul. Then the design is obtained by adding the elements of the modul to the following four initial blocks, keeping the suffixes invariant:

$(00_1, 02_1, 10_1, 12_1, 20_1, 22_1, 01_2, 12_2, 20_2, 00_3, 01_3, 02_3, 00_4, 12_4, 21_4)$
$(00_1, 12_1, 21_1, 00_2, 02_2, 10_2, 12_2, 20_2, 22_2, 01_3, 12_3, 20_3, 00_4, 01_4, 02_4)$
$(00_1, 01_1, 02_1, 00_2, 12_2, 21_2, 00_3, 02_3, 10_3, 12_3, 20_3, 22_3, 01_4, 12_4, 20_4)$
$(01_1, 12_1, 20_1, 00_2, 01_2, 02_2, 00_3, 12_3, 21_3, 00_4, 02_4, 10_4, 12_4, 20_4, 22_4)$

The author acknowledges, with thanks, suggestions made by the referee which have been incorporated in this note.

REFERENCES

Bose, R. C. (1939). On the construction of balanced incomplete block designs. *Ann. Eugen., Lond.*, **9**, 353–99.

Fisher, R. A. & Yates, F. (1953). *Statistical Tables for Biological, Agricultural and Medical Research*, 4th ed. Edinburgh: Oliver and Boyd.

Sequentially determined confidence intervals

By N. L. JOHNSON

University College, London

1. A number of authors (see Anscombe (1953) for a list of references, and in particular Stein & Wald (1947) and Wolfowitz (1950)) have shown that in many special cases the sequential method of approach does not lead to any marked advantage in problems of estimation, as compared with standard fixed sample procedures. It is the purpose of this note to demonstrate, in a simple manner, that this is generally the case under moderately simple conditions.

Suppose that a sequential procedure leading to the construction of a confidence interval for a parameter θ has the following properties:

(i) the probability that it terminates after n stages is P_n,

(ii) the conditional probability of obtaining a correct confidence interval statement, given that the sample terminates after n stages, is ω_n, and ω_n *is independent of* θ.

We define an optimum procedure as a procedure such that

(*a*) the overall confidence coefficient $\sum_{n=1}^{\infty} \omega_n P_n$ is equal to Ω,

(*b*) subject to (*a*) the average sample number $\sum_{n=1}^{\infty} nP_n$ is a minimum.

The two theorems which follow show that under fairly general conditions effectively optimum procedures need have at most two non-zero P_n's, and that, if certain further conditions are satisfied, either

(*a*) the two non-zero P's are P_m and P_{m+1}, where $\omega_m < \Omega < \omega_{m+1}$ are the nearest values of the ω's to Ω, or

(*b*) if there is an $\omega_m = \Omega$, then the procedure with $P_m = 1$ is optimum. This means that optimum procedures, as defined above, are either fixed sample procedures, or close approximations thereto.

The problem, and its solution, is similar to problems arising in the theory of linear programming, in particular those described by Beale (1955).

2. THEOREM 1. *If* $P_1, P_2, P_3, \ldots$ *satisfy the conditions*

$$\text{(i)} \quad P_n \geqslant 0,$$

$$\text{(ii)} \quad \sum_{n=1}^{\infty} P_n = 1,$$

$$\text{(iii)} \quad \sum_{n=1}^{\infty} \omega_n P_n = \Omega \quad (\omega_n > 0,\ \Omega > 0)$$

and $b_1, b_2, b_3, \ldots$ *are finite real numbers then there is a set* $\{P_n\}$ *containing not more than two non-zero values, for which either*

(*a*) $B = \sum_{n=1}^{\infty} b_n P_n$ *takes its minimum possible value, or*

(*b*) B *takes a value exceeding its lower bound by an arbitrarily small amount.*

Proof. We first prove that if B attains its minimum value for a set with a finite number of non-zero P's this number need not be greater than two.

Suppose that a finite number N ($\geqslant 3$) of non-zero P's, $P_{\alpha_1}, P_{\alpha_2}, \ldots, P_{\alpha_N}$ say, satisfy conditions (i)–(iii) of the theorem. Then values of the ratios $\delta_{\alpha_1} : \delta_{\alpha_2} : \ldots : \delta_{\alpha_N}$ can be found satisfying the conditions

$$\sum_{i=1}^{N} \delta_{\alpha_i} = \sum_{i=1}^{N} \omega_{\alpha_i} \delta_{\alpha_i} = 0. \tag{1}$$

If we define $\delta_j = 0$ for all j not equal to any α_i, then the set $\{P_n + \delta_n\}$ will satisfy conditions (i)–(iii) provided $\min (P_{\alpha_i} + \delta_{\alpha_i}) \geqslant 0$. By a suitable choice of signs of the δ's we can ensure that

$$\sum_{i=1}^{N} b_{\alpha_i}(P_{\alpha_i} + \delta_{\alpha_i}) \leqslant \sum_{i=1}^{N} b_{\alpha_i} P_{\alpha_i}.$$

Hence, increasing the numerical values of the δ's until $\min (P_{\alpha_i} + \delta_{\alpha_i}) = 0$, we obtain a fresh set of P's satisfying (i)–(iii) with N' ($< N$) non-zero values $P_{\beta_1}, P_{\beta_2}, \ldots, P_{\beta_{N'}}$, and

$$\sum_{i=1}^{N'} b_{\beta_i} P_{\beta_i} \leqslant \sum_{i=1}^{N} b_{\alpha_i} P_{\alpha_i}.$$

This argument can be repeated as long as $N \geqslant 3$. Since we start with a finite number of non-zero P's, the case $N = 3$ will eventually be reached, and a further application of the argument will reduce the number of non-zero P's to two.

Hence if B has a minimum value attained with a finite number of non-zero P's this number need not be greater than two.

Now consider sets of values $\{P_n\}$ containing an infinity of non-zero P's. Since the P's must satisfy conditions (i) and (ii) the value of B for each set must be the limit of the values of this sum for a sequence of sets with finite (though increasing) numbers of non-zero P's.

The value of B for each of these sets cannot be less than the value for some set with at most two non-zero P's. Hence it is always possible to find a set $\{P_n\}$ with at most two non-zero P's for which B exceeds its lower bound by an arbitrarily small amount.

Corollary. If, among the values of B for all sets $\{P_n\}$ with not more than two non-zero P's there is a minimum value, then this is the minimum value of B for all sets of $\{P_n\}$ satisfying (i)–(iii).

3. THEOREM 2. *If* (i) b_n *and* ω_n *are each increasing functions of* n *and* (ii) *the curve with parametric equation* (b_n, ω_n) *is convex from above, i.e. if* $n < n' < n''$, *then*

$$(b_{n''} - b_n)\,\omega_{n'} > (b_{n''} - b_{n'})\,\omega_n + (b_{n'} - b_n)\,\omega_{n''}, \tag{2}$$

then there is a set $\{P_n\}$ *satisfying conditions* (i)–(iii) *of Theorem* 1 *which minimizes* B, *and this is the set containing either*

(*a*) *only the two non-zero values* P_m, P_{m+1} *where* m *is defined by* $\omega_m < \Omega < \omega_{m+1}$, *or*

(*b*) *the single non-zero value* $P_m = 1$ *if there is an* $\omega_m = \Omega$.

Proof. This theorem is suggested by intuitive geometrical considerations, but the proof given below is perhaps more satisfying.

(i) Assume that there is no ω_m equal to Ω, but that $\omega_m < \Omega < \omega_{m+1}$.

Consider the set $\{P_n\}$ containing just two non-zero values P_α, P_β with $\alpha < \beta$. These values must satisfy the equations

$$\left.\begin{aligned} P_\alpha + P_\beta &= 1, \\ \omega_\alpha P_\alpha + \omega_\beta P_\beta &= \Omega. \end{aligned}\right\} \tag{3}$$

Hence $$P_\alpha = \frac{\omega_\beta - \Omega}{\omega_\beta - \omega_\alpha}, \quad P_\beta = \frac{\Omega - \omega_\alpha}{\omega_\beta - \omega_\alpha}$$

and $$b_\alpha P_\alpha + b_\beta P_\beta = \frac{1}{\omega_\beta - \omega_\alpha}[b_\alpha(\omega_\beta - \Omega) + b_\beta(\Omega - \omega_\alpha)].$$

Evidently we must have $$\omega_\alpha < \Omega < \omega_\beta,$$

i.e. $$\alpha \leqslant m < m+1 \leqslant \beta.$$

If $\beta > m+1$ we can replace β by β' with

$$\beta > \beta' \geqslant m+1.$$

The new values P'_α, $P'_{\beta'}$ must satisfy equations analogous to (3). Hence

$$b_\alpha P'_\alpha + b_{\beta'} P'_{\beta'} = \frac{1}{\omega_{\beta'} - \omega_\alpha}[b_\alpha(\omega_{\beta'} - \Omega) + b_{\beta'}(\Omega - \omega_\alpha)]$$

and $$b_\alpha P'_\alpha + b_{\beta'} P'_{\beta'} - b_\alpha P_\alpha - b_\beta P_\beta = \frac{\Omega - \omega_\alpha}{(\omega_\beta - \omega_\alpha)(\omega_{\beta'} - \omega_\alpha)}[(b_\beta - b_{\beta'})\omega_\alpha + (b_{\beta'} - b_\alpha)\omega_\beta - (b_\beta - b_\alpha)\omega_{\beta'}]$$

$$< 0,$$

by reason of the convexity condition (2), since $\beta > \beta' > \alpha$.

Hence B is decreased by replacing β by $\beta' < \beta$. Similarly, if $\alpha < m$ B is decreased by replacing α by α' where $\alpha < \alpha' \leqslant m$. Hence the value of B is a minimum for the set $\{P_m\}$ with the two non-zero values P_m, P_{m+1}, and the value of this minimum is

$$\frac{1}{\omega_{m+1} - \omega_m}[b_m(\omega_{m+1} - \Omega) + b_{m+1}(\Omega - \omega_m)].$$

From the corollary to Theorem 1 this is also the minimum value of B for all sets $\{P_n\}$ satisfying (i)–(iii) of Theorem 1.

(ii) If $\omega_m = \Omega$ a similar argument leads to the conclusion that the minimum value of B, among all sets $\{P_n\}$ satisfying (i)–(iii) of Theorem 1 with just two non-zero values, is attained by the set with P_{m-1}, P_{m+1} not equal to zero. The value of B for this set is

$$\frac{1}{\omega_{m+1} - \omega_{m-1}}[b_{m-1}(\omega_{m+1} - \Omega) + b_{m+1}(\Omega - \omega_{m-1})]. \tag{4}$$

But the value of B for the set $\{P_n\}$ with $P_m = 1$ is b_m, which is less than (4) by the convexity property (2).

Hence if $\omega_m = \Omega$, the set $P_m = 1$ minimizes the value of B.

4. Putting $b_n = n$ in Theorem 2, it follows that if the curve (n, ω_n) is convex from above (as defined by (2)) then the optimum sequential procedure for constructing confidence intervals with assigned confidence coefficient Ω is a fixed sample procedure or a close approximation thereto. This covers a large number of cases occurring in statistical theory, including, for example, all cases where ω_n is proportional to

$$\frac{1}{\sqrt{(2\pi)}}\int_0^{c\sqrt{n}} e^{-\frac{1}{2}t^2} dt \quad \text{(see also Anscombe, 1949).}$$

Even if the curve (n, ω_n) is not convex from above, Theorem 1 shows that there will be effectively optimum procedures with only two sample sizes, although these may now differ by more than unity.

The results obtained in this note are valid for any decision procedure (i.e. not only for the construction of confidence intervals) provided the ω_n's depend only on n and not on the values of the population parameters or the procedure by which n is chosen. For this reason they do not apply, for example, to the construction of sequential tests comparing two hypotheses, nor would they apply to the construction of confidence intervals if ω_n were allowed to depend on the parameter θ being estimated (subject to the condition $\Sigma\omega_n P_n = \Omega$ independent of θ).

REFERENCES

ANSCOMBE, F. J. (1953). Sequential estimation. *J. R. Statist. Soc.* B, **15**, 1–29.

ANSCOMBE, F. J. (1949). Large-sample theory of sequential estimation. *Biometrika*, **36**, 455–8.

BEALE, E. M. L. (1955). On the minimization of convex functions subject to linear inequalities. *J. R. Statist. Soc.* B, **17**, 173–84.

STEIN, C. M. & WALD, A. (1947). Sequential confidence intervals for the mean of a normal distribution with known variance. *Ann. Math. Statist.* **18**, 427–33.

WOLFOWITZ, J. (1950). Minimax estimates of the mean of a Normal distribution with known variance. *Ann. Math. Statist.* **21**, 218–30.

Estimation of means of normal populations from observed minima

By H. A. DAVID
University of Melbourne

1. Introduction

In life tests it is often advantageous to stop the experiment after some preassigned number of failures, say k, has been observed among m objects under test, and there are other situations where it is either impossible or impracticable to obtain more than the k smallest readings. When the characteristic measured follows an exponential distribution the problems raised are fairly simple and have been studied in detail (see, for example, Epstein & Sobel, 1953). For a normal population, to which we confine consideration in this paper, Gupta (1952) has shown how maximum likelihood as well as linear estimators of the mean μ and the standard deviation σ may be found, provided $k \geqslant 2$.

However, it may happen that only the smallest observation can ever be obtained, due to an inherent breakdown of the system after one failure. In such a case several experiments will evidently be needed for the estimation of μ and σ. An example, which has led to the present investigation, is the failure due to fatigue of an aeroplane wing which is subjected to repeated cycles of a load under which it eventually breaks, at random, in one of two positions, A or B. Here the problem is to estimate the breaking strengths, μ_A and μ_B, at A and B (expressed as the logarithm of the number of cycles to failure) from values observed when n similar wings are tested; the n breaking points are, of course, also known. It will be seen that this example introduces an additional feature in that μ_A and μ_B cannot generally be assumed equal. Before dealing with it we therefore consider the case when there are several possible breaking points which, for reasons of symmetry, may be taken to be of equal strength. This is identical with the problem of estimating μ and σ from n observations each of which is the minimum of m independent $N(\mu, \sigma^2)$ variates.

2. The symmetrical case

Let x_i $(i = 1, 2, \ldots, n)$ denote the smallest of m independent $N(\mu, \sigma^2)$ variates. Assuming the x_i to be independent we shall consider the estimation of μ and σ both by the method of maximum likelihood and the method of moments.

The likelihood function $L(\mu, \sigma)$ is given by

$$L(\mu, \sigma) = \prod_{i=1}^{n} \left\{ \frac{m}{\sqrt{(2\pi)}\,\sigma} \exp\left[-\frac{1}{2\sigma^2}(x_i-\mu)^2 \right] \left(\int_{x_i}^{\infty} \frac{1}{\sqrt{(2\pi)}\,\sigma} \exp\left[-\frac{1}{2\sigma^2}(y-\mu)^2 \right] dy \right)^{m-1} \right\}.$$

Let

$$u_i = (x_i-\mu)/\sigma, \quad z(u_i) = \frac{1}{\sqrt{(2\pi)}} e^{-\frac{1}{2}u_i^2},$$

$$Q(u_i) = \int_{u_i}^{\infty} z(t)\,dt \quad \text{and} \quad A_i = z(u_i)/Q(u_i). \tag{1}$$

Then $\log L$ may be written

$$\log L = C - n \log \sigma - \frac{1}{2} \sum_{i=1}^{n} u_i^2 + (m-1) \sum_{i=1}^{n} \log Q(u_i),$$

where C is a constant. The maximum-likelihood equations are

$$\sigma\, \partial \log L/\partial \mu = \Sigma u_i + (m-1)\,\Sigma A_i = 0 \tag{2}$$

and

$$\sigma\, \partial \log L/\partial \sigma = -n + \Sigma u_i^2 + (m-1)\,\Sigma A_i u_i = 0. \tag{3}$$

These equations may be solved numerically for μ and σ with the help of tables of the A_i (K. Pearson, 1931, Table II; also Editorial, 1955, *Biometrika*, 43, 217).

However, for n large it is much easier to estimate μ and σ by the method of moments. We have

$$\mathscr{E}(x_i) = \mu - \sigma\xi_m, \quad \operatorname{var} x_i = \sigma^2 V_m,$$

where ξ_m, V_m are respectively the mean and variance of the *largest* of m unit normal deviates. Unbiased moment estimators are therefore given by

$$\sigma^{*2} = \Sigma(x_i-\bar{x})^2/[(n-1)\,V_m]$$

and

$$\mu^* = \bar{x} + \xi_m \sigma^*/c_n,$$

where $\bar{x} = \Sigma x_i/n$ and $c_n = 1 - \frac{1}{8}(\beta_2 - 1)/n + \ldots$; c_n is the factor making σ^* an unbiased estimator of σ while β_2 is the coefficient of kurtosis of the extreme. ξ_m and V_m as well as the coefficients β_1, β_2 of the largest of m normal deviates have been tabulated by Ruben (1954) for $m \leqslant 50$.

We now compare the large sample variances of μ^* and σ^* with those of the maximum-likelihood estimators $\hat{\mu}$ and $\hat{\sigma}$. The latter are, of course, only asymptotically unbiased. From (2) and (3) it follows that

$$-\frac{\sigma^2}{n}\frac{\partial^2 \log L}{\partial \mu^2} = 1 + \frac{m-1}{n}\Sigma A_i(A_i - u_i),$$

$$-\frac{\sigma^2}{n}\frac{\partial^2 \log L}{\partial \mu\, \partial \sigma} = \frac{2}{n}\Sigma u_i + \frac{m-1}{n}\Sigma A_i(1 + A_i u_i - u_i^2)$$

and

$$-\frac{\sigma^2}{n}\frac{\partial^2 \log L}{\partial \sigma^2} = -1 + \frac{3}{n}\Sigma u_i^2 + \frac{m-1}{n}\Sigma A_i(A_i u_i^2 + 2u_i - u_i^3).$$

Table 1. *Large sample variances in terms of σ^2/n of maximum-likelihood and moment estimators ($\hat{\mu}$, $\hat{\sigma}$ and μ^*, σ^*) derived from n minima of m independent $N(\mu, \sigma^2)$ variates*

m	var $\hat{\mu}$	var μ^*	var $\hat{\sigma}$	var σ^*
1	1·000	1·000	0·500	0·500
2	0·781	0·782	0·509	0·515
3	0·796	0·806	0·514	0·532
4	0·854	0·874	0·517	0·541
5	0·922	0·956	0·520	0·550
6	0·991	1·041	0·522	0·558
7	1·057	1·124	0·523	0·566
8	1·120	1·204	0·525	0·572
9	1·180	1·281	0·526	0·578
10	1·236	1·354	0·527	0·583
12	1·340	1·491	0·529	0·592
15	1·477	1·675	0·531	0·603
20	1·668	1·937	0·533	0·617
30	1·960	2·350	0·537	0·637
40	2·183	2·672	0·539	0·651
50	2·362	2·935	0·541	0·661

To obtain expected values we require terms of the type

$$\mathscr{E}(u^r A) \quad \text{and} \quad \mathscr{E}(u^r A^2).$$

Since the density function of u is

$$f(u) = mz(u)\, Q(u)^{m-1},$$

we have

$$\begin{aligned}\mathscr{E}(u^r A) &= m\int_{-\infty}^{\infty} u^r z.zQ^{m-2}du \\ &= m\int_{-\infty}^{\infty} \frac{d}{du}(u^r z)\frac{Q^{m-1}}{m-1}du \\ &= \frac{m}{m-1}\int_{-\infty}^{\infty} (ru^{r-1}z - u^{r+1}z)\, Q^{m-1}du \\ &= \frac{1}{m-1}[r\mathscr{E}(u^{r-1}) - \mathscr{E}(u^{r+1})].\end{aligned}$$

Similarly, by two integrations by parts, we find for $m \geqslant 2$

$$\mathscr{E}(u^r A^2) = \frac{1}{(m-1)(m-2)}[r(r-1)\,\mathscr{E}(u^{r-2}) - (3r+2)\,\mathscr{E}(u^r) + 2\mathscr{E}(u^{r+2})].$$

Thus in all these cases the covariance matrix of $\hat{\mu}$ and $\hat{\sigma}$ can be obtained in terms of the first four moments of the extreme. For $m = 2$ numerical integration is necessary for the evaluation of $\mathscr{E}(u^r A^2)$.

The variances for large n of μ^*, σ^* can be obtained by standard methods. We find

$$\operatorname{var}\mu^* = \frac{\sigma^2}{n}[V_m + \tfrac{1}{4}\xi_m^2(\beta_2 - 1) - \xi_m\sqrt{(V_m\beta_1)}],$$

$$\operatorname{var}\sigma^* = \frac{\sigma^2}{4n}(\beta_2 - 1).$$

In Table 1 the large sample variances of $\hat{\mu}$, μ^* and $\hat{\sigma}$, σ^* are compared for $m = 2\,(1)\,10, 12, 15, 20, 30, 40, 50$.

3. Case of two unequal means

In this section we are concerned with the estimation of the means μ_x and μ_y of two normal populations with common variance σ^2 from a knowledge of the smaller of each of n pairs of observations x_t and y_t $(t = 1, 2, \ldots, n)$. It will be assumed that all x's and y's are independent and in addition that the population from which any observed minimum comes can be identified. Suppose that $k\ (<n)$ minima are x's and $n-k$ are y's. These observations may be regarded as the outcome of mutual *stochastic censoring*. With a slight change in notation we denote them by x_i $(i = 1, 2, \ldots, k)$ and y_j $(j = 1, 2, \ldots, n-k)$. The logarithm of the likelihood function is

$$\log L = C - n\log\sigma - \tfrac{1}{2}(\Sigma u_i^2 + \Sigma v_j^2) + \Sigma\log Q(u_i') + \Sigma\log Q(v_j'),$$

where

$$u_i = (x_i - \mu_x)/\sigma, \quad v_j = (y_j - \mu_y)/\sigma,$$
$$u_i' = (x_i - \mu_y)/\sigma, \quad v_j' = (y_j - \mu_x)/\sigma,$$

and $Q(u_i')$ is defined as in (1). With $A_i = z(u_i')/Q(u_i')$ and $B_j = z(v_j')/Q(v_j')$ the likelihood equations may be written

$$\hat{\mu}_x = \bar{x} + \frac{\hat{\sigma}}{k}\Sigma B_j, \tag{4}$$

$$\hat{\mu}_y = \bar{y} + \frac{\hat{\sigma}}{n-k}\Sigma A_i, \tag{5}$$

and

$$n = \frac{1}{\hat{\sigma}^2}[\Sigma(x_i - \bar{x})^2 + k(\bar{x} - \hat{\mu}_x)^2 + \Sigma(y_j - \bar{y})^2 + (n-k)(\bar{y} - \hat{\mu}_y)^2] + \Sigma A_i u_i' + \Sigma B_j v_j'. \tag{6}$$

For n not too large these equations may again be solved numerically.

In this case the moment approach is not so simple as in the preceding section. We note that $\bar{x}$ and $\bar{y}$ are estimators of $\mathscr{E}(x \mid x<y)$ and $\mathscr{E}(y \mid y<x)$, and therefore evaluate these two expectations in terms of the three parameters.

We have

$$f(x \mid x<y) = \int_x^\infty f(x,y)\,dy/\Pr(x<y)$$
$$= f(x)\int_x^\infty f(y)\,dy/Q(\xi),$$

where $\xi = (\mu_x - \mu_y)/(\sqrt{2}\sigma)$. It may be observed that if $\mu_x = \mu_y$ then

$$f(x \mid x<y) = 2f(x)\int_x^\infty f(y)\,dy,$$

which is just the density function of the smaller of two identically distributed variates.

Now

$$Q(\xi)\,\mathscr{E}(x \mid x<y) = \int_{-\infty}^\infty xf(x)\int_x^\infty f(y)\,dy\,dx$$
$$= \int_{-\infty}^\infty f(y)\int_{-\infty}^y xf(x)\,dx\,dy$$
$$= \int_{-\infty}^\infty \frac{1}{\sqrt{(2\pi)}\,\sigma}\exp\left[-\frac{1}{2}\left(\frac{y-\mu_y}{\sigma}\right)^2\right]\int_{-\infty}^{(y-\mu_x)/\sigma}\frac{1}{\sqrt{(2\pi)}}(\sigma z + \mu_x)\,e^{-\frac{1}{2}z^2}\,dz\,dy \quad (z = (x-\mu_x)/\sigma)$$
$$= \int_{-\infty}^\infty \frac{1}{\sqrt{(2\pi)}\,\sigma}\exp\left[-\frac{1}{2}\left(\frac{y-\mu_y}{\sigma}\right)^2\right]\left\{-\frac{\sigma}{\sqrt{(2\pi)}}\exp\left[-\frac{1}{2}\left(\frac{y-\mu_x}{\sigma}\right)^2\right] + \mu_x\left[1 - Q\left(\frac{y-\mu_x}{\sigma}\right)\right]\right\}dy$$
$$= \mu_x Q(\xi) - \frac{\sigma}{2\sqrt{\pi}}\exp\left[-\frac{1}{4\sigma^2}(\mu_x - \mu_y)^2\right].$$

Writing A_ξ for $z(\xi)/Q(\xi)$ we have

$$\mathscr{E}(x \mid x<y) = \mu_x - \frac{\sigma}{\sqrt{2}} A_\xi, \tag{7}$$

and by symmetry

$$\mathscr{E}(y \mid y<x) = \mu_y - \frac{\sigma}{\sqrt{2}} A_{-\xi}. \tag{8}$$

Similarly, we find

$$\operatorname{var}(x \mid x<y) = \sigma^2(1 + \tfrac{1}{2}\xi A_\xi - \tfrac{1}{2}A_\xi^2) \tag{9}$$

and

$$\operatorname{var}(y \mid y<x) = \sigma^2(1 - \tfrac{1}{2}\xi A_{-\xi} - \tfrac{1}{2}A_{-\xi}^2). \tag{10}$$

Equations (9) and (10) could be combined to provide a third estimating equation. However, it seems preferable and simpler to take instead k/n as an estimator of $\Pr(x<y)$, i.e.

$$Q(\xi^*) = k/n, \tag{11}$$

which gives ξ^* immediately and hence σ^*, μ_x^*, μ_y^* from

$$\bar{x} = \mu_x^* - \frac{\sigma}{\sqrt{2}} A_{\xi^*} \tag{7'}$$

and

$$\bar{y} = \mu_y^* - \frac{\sigma}{\sqrt{2}} A_{-\xi^*}. \tag{8'}$$

There is one point worth noting here. From (7′) and (8′) we have an equation for σ^*, viz.

$$\sqrt{2}\,(\bar{x}-\bar{y})/\sigma^* = 2\xi^* - (A_{\xi^*} - A_{-\xi^*}). \tag{12}$$

The right-hand side can be shown to have the same sign as ξ^*. Therefore, if ξ^* is positive (i.e. $\mu_x^* > \mu_y^*$) but $\bar{x}$ is less than $\bar{y}$, then (12) will have no satisfactory solution; likewise if $\xi^* < 0$ but $\bar{x} > \bar{y}$. This situation may arise in small samples or if μ_x and μ_y are not very different, and suggests that the simpler model $\mu_x = \mu_y$ is adequate. Even without finding such a contradiction we may, especially in large samples, decide in favour of the simpler model if k/n does not differ significantly from $\frac{1}{2}$.

4. A numerical example

Sixteen similar aeroplane wings were in turn subjected to repeated applications of the same load until breaking in one of two positions, A or B, five breaking at A and eleven at B. A and B were so far apart that the strength at A and B may be regarded as independent. Let x_i and y_j denote the *observed* breaking-strengths at A and B respectively. The following values (measured in logarithmic units from a suitable origin) were obtained:

x_i: 0·1959, 0·3856, 0·1647, 0·3655, 0·3410;
y_j: 0·2519, −0·0667, 0·2842, 0·0380, 0·5197, 0·2142, 0·2175, 0·3474, 0·3674, 0·0789, 0·1642.

Equations (4)–(6) give the maximum-likelihood estimates of μ_x, μ_y and σ as

$$\hat{\mu}_x = 0{\cdot}421, \quad \hat{\mu}_y = 0{\cdot}284 \quad \text{and} \quad \hat{\sigma} = 0{\cdot}166.$$

The corresponding moment estimates are from (11), (12), (7′) and (8′)

$$\mu_x^* = 0{\cdot}514, \quad \mu_y^* = 0{\cdot}321 \quad \text{and} \quad \sigma^* = 0{\cdot}279.$$

The sample size $n = 16$ is in this case not really sufficiently large for the simple method of estimation by moments to be altogether satisfactory, especially for σ. However, the estimation of σ was of subsidiary interest as the purpose of the experiment was to compare the estimates of μ_x and μ_y obtained from tests on wings with the results of fatigue tests on different specimens.

I am indebted to Mr D. G. Ford, of the Aeronautical Research Laboratories, Victoria, for referring to me the problem treated in §§ 3–4. Further details on applications are given in a report by Ford & Payne (1957). Dr E. J. Williams and Prof. E. S. Pearson made some valuable comments on an earlier version of this paper. Table 1 was computed by Miss B. Laby and Miss G. E. Jacobs of the University of Melbourne.

REFERENCES

EPSTEIN, B. & SOBEL, M. (1953). Life testing. *J. Amer. Statist. Ass.* **49**, 486.

FORD, D. G. & PAYNE, A. O. (1957). The fatigue characteristics of a riveted 24 S-T Wing. Part IV. Analysis of results. *Aeronautical Research Laboratories* (*Victoria*) *Structures Report.*

GUPTA, A. K. (1952). Estimation of the mean and standard deviation of a normal population from a censored sample. *Biometrika*, **39**, 260.

PEARSON, K. (1931). *Tables for Statisticians and Biometricians*, Part 2. Cambridge University Press.

RUBEN, H. (1954). On the moments of order statistics in samples from normal populations. *Biometrika*, **41**, 200.

The power of the Poisson index of dispersion

BY J. H. DARWIN

Department of Scientific and Industrial Research, New Zealand

1. The distribution of the index of dispersion, $Z = \sum_j (x_j - \bar{x})^2/\bar{x}$, for a sample of independent variables $x_1, \ldots, x_n$, drawn from a Poisson distribution with a high mean, is known to be approximately that of a χ^2 with $n-1$ degrees of freedom, say χ^2_{n-1}. It is the object of this note to find the limiting distribution of Z when the parent distribution of the x's is not Poisson but has a high mean μ, a variance of order μ and cumulants κ_r which are $o(\mu^{\frac{1}{2}r})$ for $r > 2$.

This problem, which arises in considering the power of the χ^2_{n-1} test of Z against the alternatives described above, has been tackled in part by Bateman (1950) and Kathirgamatamby (1953). Bateman discusses Neyman's contagious distribution as an alternative. She finds the first four moments of Z and shows that for a large sample size or a large mean these moments tend in form to those of $(1+m_2)\chi^2_{n-1}$, where $1+m_2$ is the ratio (variance)/(mean) for Neyman's distribution. Kathirgamatamby discusses the variable $u = \Sigma(x_j - \bar{x})^2 - (\Sigma x)\chi^2_{n-1,\alpha}/n$, where $\chi^2_{n-1,\alpha}$ is such that $\Pr(\chi^2_{n-1} > \chi^2_{n-1,\alpha}) = \alpha$. He finds the first four cumulants of u and fits distributions to these when the alternatives to the Poisson are Thomas's distribution, the negative binomial and Neyman's contagious distribution of Type A.

We shall prove that Bateman's particular result holds for the class of alternatives shown above, and we shall find the accuracy of the resultant expression for the power of Z to the next order in $1/\mu$.

2. Z may be written $(\Sigma(x_j - \bar{x})^2/\mu)/(\bar{x}/\mu)$. Suppose the distribution function of x_j has cumulants $\kappa_1 (= \mu), \kappa_2, \ldots$. Then the joint characteristic function of $x_j - \bar{x}$ $(j = 1, \ldots, n)$ is

$$E \exp[\Sigma i\theta_j (x_j - \bar{x})] = E \exp[i\Sigma x_j(\theta_j - \theta)].$$

The joint cumulant generating function of the $(x_j - \bar{x})/\sqrt{\mu}$ is then

$$\frac{\mu i \Sigma(\theta_j - \bar{\theta})}{\mu^{\frac{1}{2}}} + \frac{\kappa_2 i^2}{\mu 2!}\sum_j (\theta_j - \bar{\theta})^2 + \ldots + \frac{\kappa_r i^r}{\mu^{\frac{1}{2}r} r!}\sum_j (\theta_j - \bar{\theta})^r + \ldots. \tag{1}$$

Hence if $\kappa_r = o(\mu^{\frac{1}{2}r})$ for $r = 3, 4, \ldots$, this tends to $-\kappa_2 \Sigma(\theta_j - \bar{\theta})^2/(2\mu)$ the cumulant generating function of n normal variables with a variance matrix of rank $n-1$. It follows easily by the use of an orthogonal transformation of the $(x_j - \bar{x})/\sqrt{\mu}$ that the limiting distribution of $\Sigma(x_j - \bar{x})^2/\kappa_2$ is that of χ^2_{n-1}. Again, by Tchebycheff's theorem, $\bar{x}/\mu$ tends in probability to 1 as μ becomes large. Hence by a convergence theorem of Cramér's (1946, p. 254) the limiting distribution of $Z\mu/\kappa_2$ for large μ is that of χ^2_{n-1}.

This derivation holds for example for Neyman's contagious distribution of Type A, a negative binomial, and Thomas's distribution when each of these has a large mean μ and a ratio (variance)/(mean) which is $0(1)$ in μ.

If an alternative to the Poisson has a compound form with probability generating function $E_\lambda \exp[\lambda(z-1)]$, where λ has a distribution $dF(\lambda)$, then its cumulants κ_r are linear functions of those κ'_r of $dF(\lambda)$. The condition that κ_1 and κ_2 are large and of the same order and $\kappa_r/\kappa_2^{\frac{1}{2}r}$ tends to 0 as κ_1 becomes large for $r = 3, 4, \ldots$, is implied by the same condition in the κ'_r's. This same condition in the κ'_r's is then sufficient for the limiting distribution of Z to be of the form found above.

3. The discrete alternatives we are considering may in standard form be approximated to by an Edgeworth series for the sake of finding the joint sampling distribution of the mean and variance. Gayen (1949), using a method of Bartlett's (1935), found the joint distribution of $S_1 = \sum_{j=1}^{n} y_j$ and $S_2 = \sum_{j=1}^{n} (y_j - \bar{y})^2$, where the y's are drawn independently from a universe with zero mean, unit standard

deviation and all cumulant terms of higher order than those in λ_3^2 and λ_4 neglected, where $\lambda_r = \kappa_r/\kappa_2^{\frac{1}{2}r}$. From this joint distribution it is easy to calculate $\Pr(Z \geqslant \chi^2_{n-1,\alpha})$ to this order of approximation in the λ's.

For simplicity write $\chi^2_{n-1\,\alpha} = X$ and set

$$u = \sum_j (x_j - \bar{x})^2 - \bar{x}X.$$

If the observations x_j have been drawn from one of the alternative populations with mean $\kappa_1 = \mu$ and variance $\kappa_2 = a\mu$, then in terms of the standardized variables y and Gayen's S_1 and S_2

$$u = S_2 a\mu - X\mu - XS_1(a\mu)^{\frac{1}{2}}/n.$$

The power of the index of dispersion test based on Z, that is, the probability of establishing significance at the 100α % level when $a > 1$, is therefore $\Pr(u \geqslant 0)$.

Gayen's first approximation to the joint distribution of S_1 and S_2 is

$$g(S_1 S_2) = W(n-1) = \frac{e^{-\frac{1}{2}S_1^2/n}}{\sqrt{(2\pi n)}} \frac{S_2^{\frac{1}{2}(n-3)} e^{-\frac{1}{2}S_2}}{2^{\frac{1}{2}(n-1)}\,\Gamma(\frac{1}{2}n - \frac{1}{2})}. \tag{2}$$

If now we change our variables from S_1 and S_2 to S_1 and w, where

$$w = S_2 - \frac{XS_1}{n\sqrt{(a\mu)}} = \frac{u + X\mu}{a\mu},$$

and integrate out for S_1, we find for the leading term of the distribution of w

$$f(w) = \frac{w^{\frac{1}{2}(n-3)} e^{-\frac{1}{2}w}}{2^{\frac{1}{2}(n-1)}\,\Gamma(\frac{1}{2}n - \frac{1}{2})},$$

i.e. w is approximately distributed as χ^2 with $n-1$ degrees of freedom. Hence our first approximation to the power is

$$\Pr(Z \geqslant X) = \Pr(u \geqslant 0) = \Pr(w \geqslant X/a) = \Pr(\chi^2_{n-1} \geqslant \chi^2_{n-1,\alpha}/a),$$

which is the result reached in § 2 for the term of zero order in μ^{-1}. There is no term of order $\mu^{-\frac{1}{2}}$ in the distribution of w; that in μ^{-1} has contributions from the expansion of

$$S_2^{\frac{1}{2}(n-3)} e^{-\frac{1}{2}S_2} = \left\{w + \frac{XS_1}{n\sqrt{(a\mu)}}\right\}^{\frac{1}{2}(n-3)} e^{-\frac{1}{2}S_1 X/\sqrt{(a\mu)}}$$

in powers of $\mu^{-\frac{1}{2}}$ and from the appropriate terms in Gayen's coefficient of $W(n-1)$.

If for convenience we write

$$F_\nu = \Pr(\chi^2_\nu \geqslant \chi^2_{\nu,\alpha}/a),$$

easy calculations give for the power of the test based on Z to order μ^{-1}, and for $n > 5$

$$\begin{aligned}\Pr(Z \geqslant \chi^2_{n-1,\alpha}) &= \Pr(Z \geqslant X) \\ &= F_{n-1} + \frac{X^2}{8na\mu}(F_{n-1} - 2F_{n-3} + F_{n-5}) - \frac{\lambda_3(n-1)\,X}{4n\sqrt{(a\mu)}}(F_{n+1} - 2F_{n-1} + F_{n-3}) \\ &\quad + \frac{\lambda_4(n-1)^2}{8n}(F_{n+3} - 2F_{n+1} + F_{n-1}) + \frac{\lambda_3^2(n-1)\,(n-2)}{12n}(F_{n+5} - 3F_{n+3} + 3F_{n+1} - F_{n-1}).\end{aligned} \tag{3}$$

Taking the significance level as $\alpha = 0{\cdot}05$, some calculations have been made from this formula for $a = \frac{3}{2}$ and 2 for each of Thomas's and Neyman's distributions and for the negative binomial distribution. For these three distributions $\lambda_3\sqrt{\mu}$ and $\lambda_4\mu$ are functions of a only.

The λ values for the alternative distributions were as shown in Table 1, while the values for the power are given in Table 2, together with the approximations derived by Kathirgamatamby.

Table 1

Alternative	Case = 1·5		Case = 2·0	
	$\lambda_3\sqrt{\mu}$	$\lambda_4\mu$	$\lambda_3\sqrt{\mu}$	$\lambda_4\mu$
Thomas	1·47075	2·5714	1·71618	3·4496
Neyman	1·49691	2·7222	1·76777	3·7500
Negative binomial	1·63299	3·6667	2·12132	6·5000

The differences, with respect to ν, of the functions F_ν were found by the use of the formulae

$$\left.\begin{aligned} \tfrac{1}{2}(F_{\nu-2}(x) - F_\nu(x)) &= \frac{\partial F_\nu(x)}{\partial x}, \\ F_\nu(x) - 2F_{\nu-2}(x) + F_{\nu-4}(x) &= 2\left\{\frac{\partial F_{\nu-2}(x)}{\partial x} - \frac{\partial F_\nu(x)}{\partial x}\right\} \\ &= \left[\Gamma\left(\frac{\nu}{2}\right)\right]^{-1} e^{-\frac{1}{2}x}(x/2)^{\frac{1}{2}\nu-2}\left(\frac{x}{2} - \frac{\nu}{2} + 1\right). \end{aligned}\right\} \quad (4)$$

Table 2. *Approximations to the power of the index of dispersion test*
(The figures in parentheses are Kathirgamatamby's values.)

	$a = 1{\cdot}5$				$a = 2{\cdot}0$			
	$\mu = 1$	5	10	Main term	$\mu = 1$	5	10	Main term
$n = 101$								
Thomas	0·874 (0·870)	0·889 (0·888)	0·890 (0·890)	0·892	0·999 (0·999)	0·999 (1·000)	0·999 (0·999)	0·999
Neyman	0·869 (0·865)	0·888 (0·887)	0·890 (0·890)		0·999 (0·999)	0·999 (0·999)	—	
Neg. binomial	0·834 (0·834)	0·881 (0·880)	0·887 (0·886)		0·996 (0·988)	0·998 (0·997)	0·999 (0·999)	
$n = 51$								
Thomas	0·636 (0·635)	0·666 (0·667)	0·670 (0·669)	0·674	0·962 (0·956)	0·962 (0·960)	0·962	0·962
Neyman	0·629 (0·665)	0·665 (0·666)	0·669 (0·669)		—	—	—	
Neg. binomial	0·588 (0·606)	0·657 (0·659)	0·665 (0·666)		0·910 (0·898)	0·952 (0·949)	0·957 (0·955)	
$n = 20$								
Thomas	0·337	0·379	0·384	0·389	0·694	0·713	0·716	0·718
Neyman	0·333	0·378	0·383		0·682	0·711	0·714	
Neg. binomial	0·310	0·373	0·381		0·559	0·686	0·702	
$n = 10$								
Thomas	0·201	0·246	0·251	0·257	0·443	0·479	0·484	0·489
Neyman	0·198	0·245	0·251		0·431	0·477	0·483	
Neg. binomial	0·181	0·242	0·249		0·319	0·455	0·472	
$n = 6$								
Thomas	0·138	0·183	0·188	0·194	0·307	0·345	0·349	0·354
Neyman	0·135	0·182	0·188		0·296	0·342	0·348	
Neg. binomial	0·118	0·179	0·186		0·199	0·323	0·339	

For all n except 101 the differences were checked by the method of interpolation recommended in the introduction to Pearson & Hartley's tables (1954, p. 13). For $n = 101$ the main term in (3) was found from the Wilson–Hilferty normal approximation.

The table shows in its comparison of (3) with Kathirgamatamby's results that the main term

$$F_{n-1} = \Pr(\chi^2_{n-1} \geqslant \chi^2_{n-1,\alpha}/a),$$

which depends only on $a = \kappa_2/\kappa_1$ and not on the higher cumulants, is a good approximation to the power for most practical purposes when $n = 51$ and 101. It is less than 2·5 % higher than Kathirgamatamby's figure even for μ as low as 5. If more accuracy is required the second term gives an extremely good correction down to $\mu = 5$ and in most cases down to $\mu = 1$.

We have no independent calculation by which to check the accuracy of the correction when $n = 6$, 10 and 20 and μ has the same values as before, although it is to be expected that it will not be as good as for $n = 51$ and 101.

My thanks are due to Miss B. I. Harley for checking the main part of the calculations, which were in some places in need of correction.

REFERENCES

Bartlett, M. S. (1935). *Proc. Camb. Phil. Soc.* **31**, 223.
Bateman, G. I. (1950). *Biometrika*, **37**, 59.
Cramér, H. (1946). *Mathematical Methods of Statistics*. Princeton University Press.
Gayen, A. K. (1949). *Biometrika*, **36**, 353.
Kathirgamatamby, N. (1953). *Biometrika*, **40**, 225.
Pearson, E. S. & Hartley, H. O. (1954). *Biometrika Tables for Statisticians*. **1**. Cambridge University Press.

Some properties of the bivariate normal distribution considered in the form of a contingency table

By H. O. LANCASTER

School of Public Health and Tropical Medicine, Sydney

1. In considering the properties of contingency tables and bivariate frequency distributions Karl Pearson (1904) showed that if a bivariate normal distribution is classified in a two-way table the contingency and the correlation parameter are related by the expression

$$\phi^2 \equiv \chi^2/N = \rho^2/(1-\rho^2). \tag{1}$$

It is assumed that N, the number of observations, is large and that the class intervals are very narrow —the equation, in fact, represents a limiting property. For a p-way classification based on a multivariate normal distribution of dimension p Pearson also showed that

$$\phi^2 = 1/\sqrt{(RR')} - 1, \tag{2}$$

where R is the determinant of the correlation matrix R and R' is the determinant of the matrix $(2\mathbf{1} - \mathbf{R})$.

It was on the basis of these results that Pearson proposed his coefficient of contingency

$$C = \left(\frac{\phi^2}{1+\phi^2}\right)^{\frac{1}{2}}. \tag{3}$$

In the limiting normal case this is equal to the correlation parameter ρ. More generally, it may be regarded as an invariative property of a distribution, being unchanged by any arbitrary, not necessarily linear transformation of the scale of either or both variates.

2. In 1940 Fisher considered contingency tables from the point of view of discriminant analysis. Suppose that 'scores', i.e. arbitrary variate values, are assigned to the rows and also to the columns of a contingency table: what are the best scores to assign to the rows so that a linear function of them will best differentiate the classes determined by the columns, and vice versa? This turns out to be a problem in maximizing the correlation between the scores and the required correlations are those known as

'canonical' in the sense of Hotelling (1936). The work was continued and developed by Maung (1941). In particular, Maung quotes a result by Fisher which gives the observed frequency in terms of the canonical correlations; in fact, if the frequency is a_{ij} with marginal *totals* $a_{i.}$, $a_{.j}$ and total $a_{..}$, and if the canonical correlations are $R_1, R_2, \ldots, R_{m-1}$, we have

$$a_{ij} = \frac{a_{i.}a_{.j}}{a_{..}}\left\{1 + \sum_1^{m-1}(x_k y_k R_k)\right\}, \tag{4}$$

where x and y are the assigned scores corresponding to the given cell.

3. In this note I derive some further theorems in this field and link together some hitherto disconnected results. It will be shown that (3) is, in the limit, equivalent to the well-known Mehler identity or tetrachoric series:

$$(2\pi)^{-1}(1-\rho^2)^{-\frac{1}{2}}\exp\{-\tfrac{1}{2}(x^2-2\rho xy+y^2)/(1-\rho^2)\} = (2\pi)^{-1}\exp\{-\tfrac{1}{2}(x^2+y^2)\}\left\{1+\sum_1^\infty \psi_i(x)\,\psi_i(y)\,\rho^i\right\}, \tag{5}$$

where $\psi_i(x)$ is the Hermite–Tchebycheff polynomial of order i.

A maximal property of the bivariate normal distribution

4. We may consider the variables standardized so as to have unit variance and take $0<\rho<1$.

THEOREM. *Let x and y be jointly distributed in the bivariate normal distribution with correlation ρ. If now a transformation, $x' = x'(x)$, $y' = y'(y)$, is made to any new variables x' and y' such that*

$$(2\pi)^{-\frac{1}{2}}\int_{-\infty}^{\infty} x'^2 \exp -\tfrac{1}{2}x^2\,dx \quad \text{and} \quad (2\pi)^{-\frac{1}{2}}\int_{-\infty}^{\infty} y'^2 \exp -\tfrac{1}{2}y^2\,dy$$

are finite, then the correlation of the new variables is less in absolute value than ρ. That is, ρ is the maximum canonical correlation.

Under the conditions of the theorem, the new variables may be expressed in a series of standardized Hermite–Tchebycheff polynomials,

$$x' = a_0 + a_1\psi(x) + a_2\psi_2(x) + \ldots \tag{6}$$

such that

$$(2\pi)^{-\frac{1}{2}}\int_{-\infty}^{\infty}\left\{\left(x' - a_0 - \sum_1^\infty a_i\psi_i(x)\right)^2 \exp\left(-\tfrac{1}{2}x^2\right)dx\right\}$$

is arbitrarily small and $\sum_1^\infty a_i^2$ convergent. Moreover, the correlation is unaffected by either a linear change of scale or a change of origin, so that without loss of generality we can write,

$$\left.\begin{aligned} x' &= \sum_1^\infty a_i\psi_i(x), \quad \sum_1^\infty a_i^2 = 1,\\ y' &= \sum_1^\infty b_i\psi_i(y), \quad \sum_1^\infty b_i^2 = 1,\end{aligned}\right\} \tag{7}$$

where the $\psi_i(x)$ obey the orthogonality relations

$$(2\pi)^{-\frac{1}{2}}\int_{-\infty}^{\infty}\psi_i(x)\,\psi_j(x)\exp -\tfrac{1}{2}x^2\,dx = \delta_{ij}. \tag{8}$$

By a consideration of the expectation of $\exp(tx - \frac{1}{2}tx^2 + uy - \frac{1}{2}u^2)$, we find

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\psi_i(x)\,\psi_j(y)\,f(x,y)\,dx\,dy = \delta_{ij}\rho^i, \tag{9}$$

where $f(x,y)$ is the density function of the bivariate normal population.

The variance of x' and y' are easily found to be unity by the orthogonal properties of the ψ_j. We have then

$$\operatorname{corr}(x',y') = \sum_1^\infty a_i b_i \rho^i, \tag{10}$$

and this is less than $|\rho|$ unless $a_1 = b_1 = 1$. $|\rho|$ is therefore the maximum canonical correlation under a very general class of transformations. This was proved by Maung (1941) by an alternative method.

5. THEOREM. *The values to be assigned to the canonical variables corresponding to the canonical correlations in descending order are $\psi_i(x)$ and $\psi_i(y)$. The canonical correlations are the powers of ρ.*

We have already shown that the maximum canonical correlation corresponds with $\psi_1(x) = x$ and $\psi_1(y) = y$. Let us take a second set of values x'' and y'', such that

$$E(x'') = E(y'') = 0 \quad \text{and} \quad E(xx'') = E(yy'') = 0 \quad \text{and} \quad E(x'')^2 = 1 = E(y'')^2.$$

We may write

$$\left.\begin{aligned} x'' &= \sum_1^\infty c_i \psi_i(x), \\ y'' &= \sum_1^\infty d_i \psi_i(y). \end{aligned}\right\} \tag{11}$$

The conditions of the theorem again enable us to set $\sum_{i=1}^{\infty} c_i^2$ and $\sum_{d=1}^{\infty} d_i^2$ equal to unity.

Further $E(xx'') = 0$ forces c_1 to be zero and similarly d_1 is zero.

Now

$$\operatorname{corr}(x'', y'') = \sum_2^\infty c_i d_i \rho^i, \tag{12}$$

and this is maximal in absolute value only if $c_2 = d_2 = 1$ and all other $c_i = d_i = 0$.

This process can be extended by induction and proves the theorem.

COROLLARY 1. If a choice of variables can be made so that a joint bivariate normal distribution results from a contingency table, then the canonical correlations are powers of the greatest of them and the sets of canonical variables are the standardized Hermite–Tchebycheff polynomials.

COROLLARY 2. The roots of the determinantal equation usually solved in the treatment of this problem would be powers of the greatest of them but for disturbances due to sampling errors and difficulties caused by grouping.

COROLLARY 3. In the normal case the concept of minimum correlation has no validity. The lowest of the non-zero correlations will be ρ^{m-1} approximately and deviations from this value will be due again to sampling and difficulties caused by grouping.

FISHER'S IDENTITY AND MEHLER'S IDENTITY

6. Consider the identity (5) in the limiting case when the contingency table with elements $a_{ij}/a_{..}$ becomes the frequency function $f(x, y)\,dx\,dy$. We have

$$f(x,y)\,dx\,dy = \left\{1 + \sum_1^\infty \psi_i(x)\,\psi_i(y)\,\rho^i\right\}(2\pi)^{-1}\exp\{-\tfrac{1}{2}(x^2+y^2)\}\,dx\,dy, \tag{5 bis}$$

which is the Mehler series.

THE RELATIONSHIP BETWEEN χ^2 AND CORRELATION

7. In reconsidering the partition of χ^2 in a contingency table Lancaster (1953) used the Hermite–Tchebycheff polynomials to avoid the infinitely fine subdivision used by Pearson. In the notation of this paper he derived

$$\begin{aligned} \chi^2/a_{..} &= R_1^2 + R_2^2 + \ldots + R_{m-1}^2 \\ &\sim \rho^2 + \rho^4 + \ldots + \rho^{2m-2} = \rho^2(1-\rho^{2m-2})/(1-\rho^2). \end{aligned} \tag{13}$$

Evidently

$$\chi^2/a_{..} \geqslant R_1^2, \tag{14}$$

and R_1 is greater than any other observed correlation. A more general result is that in a complex contingency table

$$\chi^2/a_{..} \geqslant \sum_{i<j} r_{ij}^2, \tag{15}$$

where r_{ij} are the observed correlations between any variables. For if we partition χ^2 making use of the direct product of matrices (Lancaster, 1951) we may take the first row of each to be of the form $\sqrt{w_j}$ and the second row $x_j\sqrt{w_j}$, where x is suitably normalized, the second row being orthogonal to the first. The remaining rows may be filled in with due regard to the orthogonal conditions. In the observed case, the set of canonical variables for x, when the correlation between x and y is being maximized, may be different from the set when the correlation between x and z is being maximized. So that we cannot assert that $\chi^2/a_{..}$ is greater than the sum of the squares of the canonical correlations.

We can obtain a result analogous to (2):

$$\log(1+\phi^2) = -\tfrac{1}{2}\log(RR') \quad \text{(Pearson notation)}$$
$$= -\tfrac{1}{2}\log|1+\mathbf{P}| - \tfrac{1}{2}\log|1-\mathbf{P}|, \tag{16}$$

where $\mathbf{P}$ is the matrix with elements, $p_{ij} = r_{ij}$ for $i \neq j$, and $p_{ii} = 0$. If $\phi^2 < 1$, we may expand both sides using the expansion of Durbin & Watson (1950) on the right.

We have

$$\phi^2 - \tfrac{1}{2}\phi^4 + \tfrac{1}{3}\phi^6 \ldots = \tfrac{1}{2}\operatorname{tr}\mathbf{P}^2 + \tfrac{1}{4}\operatorname{tr}\mathbf{P}^4 \ldots,$$

and so

$$\phi^2 > \tfrac{1}{2}\operatorname{tr}\mathbf{P}^2 = \sum_{i<j} r_{ij}^2. \tag{17}$$

Even for $\phi^2 > 1$ this still holds, for

$$1+\phi^2 = \exp\{\tfrac{1}{2}\operatorname{tr}\mathbf{P}^2 + \tfrac{1}{4}\operatorname{tr}\mathbf{P}^4 \ldots\}$$
$$\geqslant 1 + \tfrac{1}{2}\operatorname{tr}\mathbf{P}^2,$$
$$\phi^2 \geqslant \sum_{i<j} r_{ij}^2. \tag{18}$$

There are various determinantal conditions on the r_{ij} derived by Pearson (1904) but it appears that they can all be summarized by saying that $(1+\mathbf{P})$ and $(1-\mathbf{P})$ of (16) are both positive definite. Pearson (1904) noted that the first is necessary for the system to be a correlation system and the second condition is necessary for ϕ^2 to have a finite limit.

Discussion

8. This paper relates some hitherto disconnected results. A maximal property of the bivariate distribution is proved. Further, in the theoretical case, we have derived the interesting result that the sets of canonical variables are the standardized Hermite–Tchebycheff polynomials and the canonical correlations the corresponding power of $|\rho|$. This result raises a doubt as to whether the common procedure of considering the roots of the determinantal equation separately can be justified if the hypothesis specifies a bivariate normal population. It would appear that once the first set of canonical variables are obtained, little further is gained by examining other sets since they differ from certain polynomial functions of the first set of canonical variables only because of sampling difficulty. Furthermore, a consideration of the hypothesis of the existence of one canonical correlation as in Williams (1952) shows that a value of the canonical correlation gives a limit to the range of the canonical variables; for the modified hypothesis is

$$p_{ij} = p_{i.}p_{.j}(1+\rho x_j y_i), \tag{19}$$

and $p_{ij} \geqslant 0$, so that

$$x_j y_i \geqslant -\rho^{-1} \tag{20}$$

for any pair of y_i and x_j. But both x and y must take some positive values and some negative values under the maximization procedure, so that (20) can only hold if neither x nor y have an infinite range. It appears that these difficulties have not been thoroughly explored as yet.

This paper is published with the permission of the Director-General of Health, Canberra.

REFERENCES

Durbin, J. & Watson, G. S. (1950). Testing for serial correlation in least squares regression. I. *Biometrika*, **37**, 409.

Fisher, R. A. (1940). The precision of discriminant functions. *Ann. Eugen., Lond.*, **10**, 422.

Fisher, R. A. (1941). Cited by Maung (1941).

Hotelling, H. (1936). Relations between two sets of variables. *Biometrika*, **28**, 321.

Lancaster, H. O. (1951). Complex contingency tables treated by the partition of χ^2. *J. R. Statist. Soc.* B, **13**, 242.

Lancaster, H. O. (1953). A reconciliation of χ^2. *Sankhyā*, **13**, 1.

Maung, K. (1941). Measurement of association in a contingency table with special reference to the pigmentation of hair and eye colours of Scottish school children. *Ann. Eugen., Lond.*, **11**, 189.

Pearson, K. (1904). Mathematical contributions to the theory of evolution. XIII. On the theory of contingency and its relation to association and normal correlation. *Drap. Co. Mem. Biom. Ser.* no. 1.

Williams, E. J. (1952). Use of scores for the analysis of association in contingency tables. *Biometrika*, **39**, 274.

REVIEWS

Einführung in die mathematische Statistik. By L. SCHMETTERER. Vienna: Springer-Verlag. 1956. Pp. xxiii + 405. £4. 3*s*. 6*d*.

In this book it is evident that mathematical rigour has had first call (indeed, the publishers say explicitly that it is primarily for mathematicians) and a level of mathematics corresponding to the English B.Sc. General is assumed in the reader, augmented by knowledge of the elements of the Stieltjes integral, elementary set theory, Borel measure of sets and the Lebesgue integral. But, for all this, the mathematics has been kept to a minimum, and the general reader who is prepared to take a few theorems on trust will find this an eminently readable book. Thus the first chapter starts with the Kolmogorovian axiomatics, developing the set theory as needed, and follows with the theory of characteristic functions omitting the proofs of the uniqueness and limiting theorems. The disadvantages of the mathematical approach have not been entirely avoided however, and the second chapter treating distribution theory (together with the associated tests at an intuitive level) begins by proving Cochran's theorem, deducing, for instance, the distribution of the normal sample variance therefrom. The third chapter gives a straightforward account of confidence intervals and the fourth gives a rigorous but concise treatment of estimation theory. The fifth chapter giving the theory of test functions has recently been recommended by Neyman in *J. R. Statist. Soc.* (It is a pity here that Wilks has not been given credit for his theorem on likelihood ratio tests and Kolodziejczyk's general linear hypothesis has not been given explicitly.) The sixth chapter develops multiple regression theory and the elements of multivariate analysis; Wishart's, Hotelling's, Fisher's discriminant and Mahalanobis's D^2 distributions are derived by elementary methods. The seventh chapter on 'non-parametric tests' is the least satisfactory, being rather scrappy and eclectic (but it only claims to be an introduction). Ordered variables, the probability integral transform, the runs-, Kolmogorov–Smirnov-, Wilcoxon- and sign-tests are briefly dealt with (Continental contributions being given here more credit than is perhaps their due) and there is little mention of power. The short eighth chapter treats briefly of the classical Bayes theory and its consequences in regard to modern theory. The whole is preceded by full German–English glossary and followed by an index of moderate length. [*Note*: the index-reference to χ^2 as on p. 379 must be a misprint for p. 279.]

Criticisms of the book are chiefly on the score of omissions and relative neglect, partly no doubt due to reasons of space and the nature of the treatment. I feel that a fuller treatment of χ^2 is deserved; contingency tables, index of dispersion, Eisenhart's theorem on the power and optimum grouping do not appear to be mentioned. More complex analysis of variance models and their powers should be at least described. Combinatorial methods get scant treatment (Mood's tests are not mentioned) and discrete variables also. I cannot find that the negative binomial, Polya's or Neyman's Contagious, distributions are mentioned, while rank correlation gets only a footnote reference. The distribution of the correlation coefficient, Smirnov's ω^2 and Neyman's ψ^2 also seem to be absent. Sequential testing could well be given a few pages rather than a bare mention.

D. E. BARTON

Statistical Methods and Scientific Inference. By SIR RONALD A. FISHER. Edinburgh: Oliver and Boyd. 1956. Pp. viii + 175. 16*s*.

This book is welcome in providing an up-to-date and authoritative account of Sir Ronald Fisher's views on statistical inference. Readers new to statistics are, however, advised not to try to read Fisher's book entirely on its own. The issues discussed are fairly technical and sometimes controversial; and this book largely concentrates on the author's original contributions, without always adequate representation of other points of view. The book is in consequence likely to repay study most by those already well acquainted with current statistical methods.

The first two chapters are mainly historical, but at the end of the second chapter we first come to the essential step of saying what is meant by probability (essential because our attitude on probability will colour our whole approach to the inference problem). A more detailed discussion would have been useful on what is meant by *statistical* inference; for the reviewer's experience is that the word 'statistics' is also interpreted differently by different people. Thus in the recent American book, *The Foundations of Statistics*, by L. J. Savage, it is implied in the introduction that mathematical statistics is to be

identified with the 'mathematical treatment of problems of inductive inference'. The historical and scientific association of the word 'statistics' with those types of observational phenomena, in which the properties of the group rather than of the individual are considered, seems there ignored. Yet it is this aspect which has given rise to the frequency or statistical interpretation of probability, in contrast with the notion of probability as a degree of belief or credibility. Fisher, while favouring the statistical interpretation, emphasizes that, before a probability defined by some aggregate or population is associated with any *particular* event, we must be unable to associate any 'subaggregate', with a different probability, to this same event. This is a relevant inductive consideration, though it seems logically distinct from the concept of statistical probability itself, and there may be difficulties in assigning to it any very precise meaning. It is well known that there are at least two rival schools of thought on statistical inference: either one can play down the statistical aspects, and argue explicitly in terms of inverse probability, prior and posterior probabilities, and perhaps utilities, or one can attempt, as many statisticians do, to summarize statistical features of the data as an aid to, but not necessarily as a substitute for, one's final inductive inferences or decisions. The impression gathered from this book is that Fisher is putting forward methods that logically lie somewhere between these two schools, but in the reviewer's opinion these methods create difficulties and apparent inconsistencies that outweigh any possible advantages. Thus in Chapter 3 (bottom of p. 42) Fisher uses such a phrase as 'a test of significance applied to a single hypothesis by a unique body of observations', but omits to explain either on what justification the concept of a random sample is applied to such unique data, or on what basis the critical region, in the Neyman–Pearson sense, of the significance test is arrived at (a small probability for the observations is not sufficient; for example, any hand at bridge has the same probability as a hand with all trumps). Both in this book and elsewhere Fisher has criticized the notion of 'repeated sampling' without appearing to recognize its close relation to the concept of the random sample.* The idea that the random variation should be restricted as far as possible, so that only samples as far as possible like the one observed should be considered, is a valuable one which Fisher has promoted on several occasions; if the statistical viewpoint is to be retained, however, there is a limit to what can be done in this way. For example, it is possible to test the difference in means of two samples either on the assumption of normality or on a permutation test based on the actual observational values; but insistence on the latter approach is neither customary nor (for very small samples) necessarily rewarding. The failure of the Behrens–Fisher test to have the frequency interpretation required by many statisticians arises in the reviewer's opinion from an inconsistent attitude to the observed variance ratio s_1^2/s_2^2, which is at one stage of the derivation considered random and at the next considered fixed.† The suggestion (p. 56) that no new axioms are needed for fiducial inference would thus not be generally accepted.

The exposition of Fisher's theory of fiducial inference given in this book is as complete as will be found anywhere, but the 'theory' when more than one unknown parameter is present still seems left rather more a matter of definition for each individual problem than a general method. In the case of one parameter, the theory was accepted by many statisticians because of its frequency interpretation, which renders it practically equivalent to the theory of (efficient) confidence intervals obtained from

* Cf. E. S. Pearson's remarks in 'Statistical concepts in their relation to reality', *J. R. Statist. Soc.* B, **17** (1955), 204–7.

† It is not the primary intention of the reviewer to answer here specific references in the book to his earlier criticisms of the Behrens–Fisher test, but the following remarks are relevant:

(i) (p. 96) The 'randomization test' was put forward purely to demonstrate its incompatibility on a frequency interpretation with the Behrens–Fisher solution. The criticisms by Fisher of the use of randomization with tests of significance have some weight, but the use of randomization in the theory of games and in experimental design are also not above criticism if attention is focused on a *single* decision or a *single* experiment. The test actually used on a practical problem in the first critical paper (*Proc. Camb. Phil. Soc.* **32** (1936), 560–6) ascribed a significance level lying between the values obtained on the assumptions $\sigma_1^2/\sigma_2^2 = 0$ and 1. A closer assessment of the significance level, taking into account the observed variance ratio s_1^2/s_2^2, is now available from Welch's test (*Biometrika*, **34** (1947), 28–35).

(ii) (pp. 120 and 162) The reviewer's original reference to the particular function $\mu + \sigma$ was unfortunate, for he had overlooked the confidence (as well as fiducial) interval solution for the particular linear function $\mu + \alpha\sigma$; this solution follows from the sampling distribution of the 'non-central t' being independent of σ. The general principle still, however, stands that a confidence interval solution for an arbitrary function of two unknown parameters does not follow automatically from a simultaneous sampling distribution involving them. Even in Fisher's fiducial theory, the problem of uniqueness is hardly settled by the discussion of a particular example on pp. 169–72.

continuous random variables. In the book confidence intervals are almost entirely referred to in relation to discrete variables, but as Fisher has rejected the frequency basis for fiducial intervals the value of confidence intervals in all cases must be considered. Such confidence intervals are admittedly of more relevance in some applications than others; however, such routine statistical methods fortunately, if not used too automatically, often have an inductive value even when the data are more 'unique'. In the case of *small* sample theory, over which there is likely to be most argument, this inductive value arises partly because statistical experience (external to the sample) on the nature of the population often exists. For fiducial inference Fisher has introduced the requirement (p. 56) that no knowledge *a priori* on the value of the parameter should be available. This is desirable if the fiducial inference is to be regarded as properly inductive, but the scope of fiducial inference seems now academically narrow, and the logical status of other prior knowledge or assumptions in the theory is left extremely obscure.

The problem of specification arises again in the section beginning on p. 66 (cf. also p. 127), where an explicit record of the likelihood function, as it varies with an unknown parametric value, is proposed as a summary of the data. The likelihood function is certainly relevant and important for many statistical inferences; nevertheless, any table proposed on the basis of a *particular probability model* would rarely be an adequate substitute for the data themselves, as the model might not be accepted as appropriate by other workers. Standard statistical methods become less available when the data or phenomena under consideration are highly individual, and standard inductions likewise. Even when the likelihood function is accepted as appropriate, the circumstance that it does not involve the number of ways by which the observed outcome might have occurred implies that it is not dependent on the sampling procedure (for example, whether direct or inverse binomial sampling), a knowledge of which is sometimes required.

Apart from the concluding section on simultaneous fiducial distributions, the last chapter covers the theory of estimation associated with the name of the author; the importance and established value of this theory might well have justified its consideration earlier, except that it is now fairly familiar to the readers who, it has been here suggested, will most profit from this book. Such readers will be entertained but not otherwise diverted by Fisher's occasionally provocative style. The problem of inductive scientific inference, and in particular statistical inference, is one in which impartiality and tolerance are as essential to further progress as rules of procedure. Any suspicion of dogmatism, from whatever quarter, or however illustrious the person concerned, would be regrettable.

M. S. BARTLETT

Symposium on Monte Carlo Methods. Edited by H. A. MEYER. New York: John Wiley and Sons, Inc.; London: Chapman and Hall, Ltd. 1956. Pp. xvi + 382. 60*s*.

Interest in Monte Carlo methods is increasing again after suffering a depression around 1952. It is reflected in this book which contains papers presented at a symposium held at the University of Florida on 16 and 17 March 1954. There are twenty contributions headed by an introduction by A. W. Marshall which is intended as a guide to the non-specialist; this traces the development of Monte Carlo since its christening in about 1947 and attempts to place the succeeding papers in perspective, giving a brief description of most of them.

Three papers are devoted to the generation and testing of random variables. O. Taussky and J. Todd give a brief survey of the available methods of generating pseudo random numbers by arithmetical processes and present the results of some tests on these. J. W. Butler's paper contains many useful ideas on the important subject of transforming these random numbers into samples from given probability distributions. He discusses the three basic methods which he calls the direct, composition and rejection methods. It should perhaps have been pointed out in discussing these that the last, using as it does a variable number of random numbers, is often inconvenient when several correlated problems have to be studied.

The remaining papers vary from the straightforward to the highly original. Of the latter that by H. F. Trotter and J. W. Tukey is outstanding. Here, amid many comments that show a very fine appreciation of the fundamental principles of good Monte Carlo practice, we are introduced to a new variance-reducing technique. This is applied to a complicated problem about normal samples in another paper where its success is amply demonstrated. It is a pity that the paper is marred by a rather obtuse explanation of the very heart of their idea.

H. Kahn gives a very useful collection of formulae and results connected with many of the now commonly used Monte Carlo tricks, for several of which he is largely responsible. This is clearly the

voice of a very widely experienced practitioner of Monte Carlo. The diffusion of γ-rays through shields is amply treated in two papers by M. J. Berger and by L. A. Beach and R. B. Theus, in which can be found examples of many of the techniques described by Kahn. On the purer mathematical side, J. H. Curtiss gives a lengthy account of an investigation into the relative efficiencies of a Monte Carlo and the normal computational procedures for solving a set of linear simultaneous equations.

The book ends with a bibliography of some 300 entries divided into four sections, the first two, on Monte Carlo proper and random number generation, containing generous abstracts. There are source, name and subject indices to the whole book.

Though most of the subject-matter in this book is not new it is a valuable collection of ideas that had previously had only a limited circulation. It is a pity it could not have been produced more closely after the date of the symposium.

K. W. MORTON

The Essentials of Educational Statistics. By F. G. Cornell. New York: John Wiley and Sons Inc.; London: Chapman and Hall Ltd. 1956. Pp. 375. 46*s*.

The author writes in his preface: 'as an introduction to statistics this book is not a *handbook*, though it contains more than enough material for the usual first two semesters. The interest has been to include *essentials* both of specific techniques and of underlying ideas. It seems reasonable that a book with this purpose should cover a limited ground thoroughly rather than much ground lightly.' In spite of the author's words, however, this book undoubtedly falls into the category of handbooks.

The student envisaged as a reader of this book must be one with no mathematical knowledge, since no derivations of any kind are given, the necessary theorems being stated quite clearly but without symbolism. For example, on pp. 113–14 when beginning to discuss the sample mean the author states as principles:

(*a*) The mean of means of all possible random samples of size n drawn from a population equals the mean of the population.

(*b*) The variance of the sampling distribution of means of size n is equal to $1/n$ times the variance of the population.

(*c*) The sampling distribution of means of samples of size n from a normal population is itself normal.

(*d*) The sampling distribution of means of size n ... for a wide variety of non-normal populations approaches a normal distribution as n becomes increasingly large.

[It might have been well to have inserted 'with replacement' or possibly 'independent' somewhere in these principles but this point is mentioned in the discussion in the text.] This method is quite clearly pursued throughout the book and may be found adequate for students who want to know how the various standard techniques work without asking why they are used and how their properties are derived. It is inevitable, however, that mathematical symbols are used for criteria, test functions etc. and this may cause the student who likes words, but no mathematics, some anxiety.

The topics covered are the sampling distribution of the mean (population σ known and not known), significance tests for correlation and regression, χ^2 for goodness of fit and in contingency tables, two sample tests, simple analysis of variance, curvilinear regression and partial and multiple correlation. Abbreviated tables of the normal curve and of t are appended.

The title itself is a misnomer. The link with education is tenuous and what the author has done is to write a very elementary statistics text-book.

F. N. DAVID

Elementary Statistical Methods (revised). By W. A. Neiswanger. New York and London: The Macmillan Company. 1956. Pp. 749. 48*s*. 6*d*.

This is a substantial text-book suitable for use in a one year's course in statistics. It is revised substantially from the 1943 edition when discussing sampling methods and is extended by chapters on sampling errors and statistical inference. In much of the book the illustrative data is more recent than was available at the time of the last edition.

The book opens with an excellent general discussion on the use of statistical methods and the interpretation of results, including an extensive section on the way to ask questions and design a sample. Then follows a series of chapters on presenting the data obtained by means of charts and graphs, frequency distributions and calculated measures of average, dispersion and skewness. There

are two chapters on sampling errors and statistical inference, two on index numbers, one of which summarizes those available in the United States, three on the analysis of time series including seasonal and cyclical variation and two on regression analysis and correlation. There are appendices on arithmetic and computational methods, and on the symbols commonly used; there are tables of random numbers, the normal integral and of values of t but not of χ^2. Each chapter is followed by a short bibliography and a set of questions.

The exposition throughout is very clear and within the field chosen most aspects of the subject are treated well and completely. In an elementary text-book there is always a problem of deciding how detailed to make the exposition and how much to leave to the student's intelligence. Too much detail makes the task before the student look worse than it is and he may learn a lot of detail before he masters general principles; but if there is too little detail he may fail to follow essential steps. In my opinion this book, if anything, errs on the side of giving too much detail. It will perhaps seem rather ponderous, especially to those students who are taking several courses in the same session, and it might be better to give the student the idea and let him gradually sort out the snags himself as he meets them. For instance, does the student really need to be told that if he makes the scale of a chart too large some of the observations may occur outside the limits intended? On the other hand, detailed descriptions of the errors which may arise in the calculation of an arithmetic mean from a frequency distribution instead of from an individual list and of the method of calculating the true mean when working from a guessed value of the mean without using algebra are very good indeed, though in the former case the problem of recorded values tending to bunch, say at multiples of ten, might have been stressed. The problem arising from assuming all values in a frequency class are concentrated at the mid-point of the class in the calculation of the standard deviation is not discussed in a similar way, and the correction factor $(-\overline{x}^2)$ in the short cut method is merely stated, not explained.

Calculations in the text were obviously undertaken with assistance from a calculating machine and products and quotients are commonly shown to seven, sometimes to ten or eleven significant figures. Early in the book there is a discussion on spurious accuracy and on errors arising in rounding, but there is little indication later of the application of these excellent precepts. Nor is guidance given to the student as to what he can do in practice to avoid having to use as many significant figures as in the text of the book if he possesses, for example, not a calculating machine but only a slide rule or the tables of logarithms and square roots at the end of the book. The method used produces some curious results. For instance in the calculation of a correlation coefficient we have

$$r = \frac{135{,}919{,}105}{176{,}439{,}387} = 0{\cdot}77,$$

whilst individual values of a trend line, fitted by the method of least squares are given to seven significant figures where in only four out of thirty-five values does the fitted line represent the original series to the second figure. Looking at the calculations for this trend line made by two methods one may well doubt if it is true that the simplicity of the short cut method strongly commends it. The student might well also wonder why the short cut method for calculating a standard deviation gives a result 28·906 cents and the method working with deviations from the true mean gives a result, on the facing page, of only 28·9 cents, and this after having a sum of deviations squared calculated to eleven significant figures.

The English student of economics using this book for an ancillary course in statistics will conclude that in America figures are not a scarce factor of production and hence go on to assume that any real short cut methods are added for academic interest.

In general this is a good book for students who have the time and are prepared to learn the elements of a subject slowly but methodically, hence getting a good grounding.

H. S. BOOKER

Statistics: a New Approach. By W. Allen Wallis and Harry V. Roberts. Illinois: The Free Press. 1956. Pp. xxxviii + 637. $6.00.

This is a cookery-book designed for those wishing to know about statistics who have no mathematical learning. The logical processes of the mathematics of statistical theory are explained in words with a bias—or so the reviewer thinks—in favour of those interested in economic applications. The book is divided into four sections entitled, the Nature of Statistics, Statistical Description, Statistical Inference and Special Topics.

Statistical Description covers univariate and bivariate frequency distributions and measures of location, dispersion and association. In *Statistical Inference* we have probability and randomness,

a summary of tests based on the normal curve, and some notes on estimation. The section of *Special Topics* is a somewhat 'mixed bag' containing the design of experiments, quality control, curvilinear regression and time series. Formulae are given freely but no derivations or justification other than in words are attempted.

At the level of exposition aimed at the book may prove useful to the non-mathematician. Those with any knowledge of mathematics for whom, according to the preface, the book is also intended to be useful, are likely to go empty away.

F. N. DAVID

Theoretical Genetics. By R. B. GOLDSCHMIDT. University of California Press (for whom Cambridge University Press act as agents). 1956. Pp. 563. 64*s*.

Many readers of *Biometrika* are interested in genetics, and some of them have at one time or another contributed to its mathematical theory. All these could profitably read the *Theoretical Genetics* of Richard Goldschmidt, the grandmaster of contemporary genetics and president of the last International Congress of the subject. They might find—and perhaps be disturbed by their discovery—that genetics is not necessarily a happy hunting ground for statisticians having but little biological knowledge. While the book does not cover the more mathematical branches of genetics, such as population genetics or the theories of mutation and selection, it provides an exhaustive and authoritative discussion of the concept of the gene, as it developed during the author's lifetime and in fact to a large extent by his own stimulating influence. The main topics are (1) the nature of the genetical material (239 pages), (2) the mode of action of the genetical material in controlling specific development (234 pages) and (3) the consequences of the nature and actions of the genetical material for evolution (18 pages).

These three topics do not of course cover the entire field of modern genetics and their treatment does not aim at completeness even within their restricted framework; but the discussion even if subjective and selective is of the greatest interest for anybody who wants some clarity in his ideas of the most central and possibly most intricate field of biology. He will find some branches of genetics, if not ramshackle, at least far more dubious and incomprehensible than he would assume from a perusal of the ordinary text-books. He may also at times be amused by the exposure of premature and rather childish generalizations by meritorious and even famous research workers in the field of biochemical genetics, or by the attempts of embryologists to keep genetics out of their field.

Goldschmidt is quite explicit in describing 'gene action' as we infer it as differential action, i.e. as resulting in abnormal or at least divergent development, but he does not draw the consequences of this fact. He does not consider genetics as a science of differences but rather as a normative science comparable to embryology or biochemistry. This seems to the reviewer the main weakness of this excellent book. Fifty pages of Bibliography and two indices greatly add to its value.

H. KALMUS

Rank Correlation Methods (second edition). By M. G. KENDALL. London: Charles Griffin and Co. Ltd. 1955. Pp. 196. 36*s*.

This is a revised and much enlarged second edition of a book which first summarized the theory of rank correlation. In its new form the book brings up to date the theory of ranking and may be considered indispensable for the research worker in both the theoretical and practical fields.

Probability (second English edition). By A. N. KOLMOGOROV. New York: Chelsea Publishing Company. 1956. Pp. 84. $2.50.

This is a second edition of the translation of the original German monograph *Grundbegriffe der Wahrscheinlichkeitsrechnung* by the world's leading probabilist. It is a book from which nearly all our modern ideas may be deemed to have sprung. In this, the second edition, there are no alterations. A supplementary bibliography by A. T. Bharucha-Reid, which neither adds to nor detracts from the value of the book, has been added in an attempt to 'reflect the present status and direction of research activity in the theory of probability'. This is possibly a mistake, since the bibliography is ephemeral while the book is part of probability.

Irrationalzahlen (second edition). By O. PERRON. New York: Chelsea Publishing Co. 1951. Pp. 199. Paper \$1.50; cloth \$3.25.

This is a reprint of the second edition of Perron's well-known work. (The third edition, of 1947, differed from the second only in one page and a footnote.) The paper and printing are good, and the reviewer could find no serious misprint.

Starting from a set of twenty-one axioms satisfied by the rational numbers, the author defines irrational numbers as Dedekind sections of the rationals, and shows that they satisfy the same axioms. He goes on to deal with limits, powers and logarithms. Then without further use of analysis he discusses various methods (particularly continued fractions) of approximating to irrational by rational numbers. The degree of accuracy attainable is investigated. He concludes with a chapter on algebraic and transcendental numbers, proving that the numbers e and π are transcendental. This means that neither of them is a root of an algebraic equation, of any degree, with integral coefficients.

The work may be regarded largely as an introduction to the author's *Die Lehre von den Kettenbrüchen*. There is a good bibliography.

G. I.. WATSON

Trigonometrical Series (second edition). By A. ZYGMUND. New York: Chelsea Publishing Co. 1952. Pp. 329. Paper \$1.50; cloth \$4.95.

It is now twenty-one years since Zygmund's Trigonometrical Series appeared in the *Monografje Matematyczne*. The present edition is of the corrected reprint of 1952. To have produced this beautifully printed volume on excellent paper at \$1.50 is a triumph of publishing economics.

There is still no book on the subject which compares with Zygmund's for clarity or thoroughness. It is a mathematician's book, but apart from a knowledge of Lebesgue integration it presupposes no specialized knowledge of mathematics. Although the character of the book is to approach questions of convergence and summability concretely, many readers will find in it an excellent introduction to more generalized theories such as that of Linear Operations. There is a long chapter on Riemann's theory of trigonometric series and a short one on Fourier's integral. Examples and miscellaneous theorems supplement the text, and there is an extensive bibliography.

No student of advanced mathematics and no mathematical library should be without this classic of analysis; the publishers are to be congratulated on their enterprise in bringing the volume within the reach of all.

H. KESTELMAN

Publications of the Mathematical Institute of the Hungarian Academy of Sciences. Vol. 1, Fasc. 1 and 2. 1956. [Subscriptions through the trade organization 'Kultúra', Sztálin út 21, Budapest VI, Hungary.]

The Mathematical Institute of the Hungarian Academy of Sciences announces that this is their new name and that the above publication continues the series 'Publications de l'Institut des Mathématiques Appliquée de l'Academie des Sciences de Hongrie', of which three volumes have been published [vol. 1 (1952), vol. 2 (1953), vol. 3 (1954)]. The change of the title of their publications is a consequence of the corresponding change of the name of the Institute on 1 August 1955 from the Institute for Applied Mathematics of the Hungarian Academy of Sciences.

Methods in Numerical Analysis. By K. L. NIELSEN. New York and London: The Macmillan Company. 1956. Pp. 382. 48*s*. 6*d*.

'Practically every university is now teaching a course in numerical analysis', says Dr Nielsen in his preface, 'and there is a need for an elementary textbook.' This purpose his book fulfils. It has nine chapters, titled I, Fundamentals; II, Finite Differences; III, Interpolation; IV, Differentiation and Integration; V, Lagrangian Formulas; VI, Ordinary Equations and Systems; VII, Differential and Difference Equations; VIII, Least Squares and their Application; IX, Periodic and Exponential

Functions. There are also answers to exercises, a bibliography, and nineteen tables, including eleven tables of coefficients for interpolation, differentiation and integration.

Securing his rear at the start with a list of explanations of symbols such as $=$, $>$ and $\surd$, Nielsen covers a good deal of ground, many topics being treated at length with mathematical proof and others (e.g. relaxation—three pages, one worked example) briefly but with references. Matrix inversion and the solution of sets of simultaneous linear equations are done exclusively by Crout's method; mention might well have been made of the requirement for adjusting the elements of the matrix by powers of 10 where their sizes are disparate. The treatment of least squares and data fitting (Chapters VIII and IX) will not serve adequately as a text for statistical computation, since it is directed solely to the evaluation of the regression constants.

To the student, a particularly valuable feature of the book is the array of clearly explained and well laid out illustrative examples. It is also a useful reference book.

T. LEWIS

PUBLICATIONS OF THE U.S. DEPARTMENT OF COMMERCE, NATIONAL BUREAU OF STANDARDS

(i) Tables of functions and zeros of functions. Applied Mathematics Series no. 37. 1954. Pp. ix + 211. $2.25.

This is a book containing sixteen divers tables of special functions and zeros of special functions, chiefly related to Bessel Functions or which have been derived as ancillary to their computation. They have almost entirely appeared previously in the *J. Math. & Phys.* and the *Bull. Amer. Math. Soc.* [1942–9]. Those in the former journal, being designed as aids to physicists, have little statistical application, but we may mention Table 6, of the Struve functions, which should prove useful in tabulating the p.d.f.'s of simple forms of non-central χ^2. Tables 7 (of Fourier coefficients), 8 (of $\sin x$ and $\cos x$ for $x = 100(1)1000$), 9 (for finding logarithms to 25 places) and 10 (of $x^n/n!$) will be occasionally useful. The tables of aids to special quadrature methods should be of frequent assistance to the research statistician however. These are Tables 11 (of 'the zeros of the Legendre polynomials of orders 1–16 and weight coefficients for Gauss' mechanical quadrature formula') and 12 (of 'zeros and weight factors of the first fifteen Laguerre polynomials'). Table 11 is, in fact, the now standard table of Lowan, Davids and Levenson. Each table is preceded by an introductory essay, with full references, especially noting articles reviewing the state of tabulation of the function and its relations.

(ii) Contributions to the solution of systems of linear equations and the determination of eigenvalues. Edited by OLGA TAUSSKY. Applied Mathematics Series no. 39. 1954. Pp. 139. $2.00.

The symposium has rather more coherence than is usual in this type of publication; many of the contributions are complementary and all are, or contain, reviews of the literature. The book lacks an index and the list of contents is so meagre as to be misleading. L. Fox gives an exhaustive description of the 'elimination' and related methods of solution of linear equations, with a wealth of worked examples, and A. I. and G. E. Forsythe discuss punched card methods for the 'accelerated gradient' method. R. M. Hayes treats the convergence of a class of iterative processes for the solution of, *inter alia*, Fredholm integral equations and multiple integrals. Other contributions discuss the 'condition' of the $n \times n$ matrix $S_n = \| (i+j-1)^{-1} \|_{i,\, j=1,\ldots,\, n}$ and table S_n^{-1} for $n = 1, 2, \ldots, 10$. Lower bounds for the rank of square matrices, location of latent roots ('eigenvalues') and bounds for them are also treated.

(iii) Table of the Descending Exponential. Applied Mathematics Series no. 46. 1955. Pp. 76. 56 cents.

The negative exponential function, e^{-x}, is tabulated to 20 decimal places at intervals of 0·001 for $x = 2{\cdot}5$ to $x = 10$.

Tables of the Cumulative Binomial Probabilities. (Ordnance Corps Pamphlet PB 111389). Washington: U.S. Department of Commerce, Office or Technical Services. 1952. Pp. viii + 577. $6.00.

Through the well-known relation

$$\sum_{r=c}^{n} {}^{n}C_r p^r (1-p)^{n-r} = I_p(c, n-c+1),$$

where I is the incomplete B-function ratio, these tables may be used to give either the cumulative binomial probabilities or incomplete B-function values. The function is tabled to 7 decimal places for

$$p = 0{\cdot}00(0{\cdot}01)\,0{\cdot}50 \quad \text{and} \quad n = 1(1)150.$$

No information is given as to the method of computation. Previous tables of this kind include:

(*a*) the National Bureau of Standards Tables (Applied Mathematics Series 6, 1950) which covered the range $n = 2(1)49$, and were derivable from Karl Pearson's *Tables of the Incomplete Beta-Function* (Cambridge University Press, 1933).

(*b*) H. G. Romig's 50–100 *Binomial Tables* (Wiley, 1953) for which $n = 50(5)100$.

The present tables, therefore, cover important fresh ground for $n > 100$.

CORRIGENDA

Biometrika (1955), **42**, pp. 531–3

'The likelihood ratio test for Markoff chains.' By I. J. Good

Mr Leo A. Goodman has been kind enough to allow me to see some work that he has not yet published in which, among other things, he points out that my paper in *Biometrika*, **42** (1955), 531–3 contains a number of inaccuracies concerning the non-cyclic (non-circularized) case. In addition he mentions the following errata.

Equation (7) should read $\bar{K}_0 = 2N \log N$, $\bar{K}_{-1} = 2N \log (Nt)$.

In the last line § 5, $\nabla\psi_\nu^2$ should be replaced by $\nabla\psi_{\nu+1}^2$.

For further details the reader is referred to Mr Goodman's forthcoming paper 'Simplified group tests for Markoff chains'.

I. J. G.

OTHER BOOKS RECEIVED

The Elements of Probability Theory and some of its Applications. By H. CRAMÉR. New York: John Wiley and Sons, Inc.; London: Chapman and Hall. 1955. Pp. 281. 56*s*.

Infinite Sequences and Series. By K. KNOPP, translated by F. BAGEMIHL. New York: Dover Publications Inc. 1956. Pp. 186. $3.50 cloth; $1.75 paper.

Numerical Analysis, vol. VI [Proc. 6th Symposium in Applied Mathematics of the American Mathematical Society held 1953]. Edited by JOHN H. CURTISS. New York, Toronto and London: McGraw Hill Book Company Inc. 1956. Pp. 303. 73*s*.

Statistical Mechanics. By T. L. HILL. New York: McGraw Hill Book Company Inc. 1956. Pp. 392 + 41 pp. appendices and index. 67*s*. 6*d*.

Nonparametric Statistics. By S. SIEGEL. New York: McGraw Hill Book Company Inc. 1956. Pp. 239 + 74 pp. references and appendix. 49*s*.

Immigrants and their Children, 1850–1950. By E. P. HUTCHINSON. New York: John Wiley and Sons Inc.; London: Chapman and Hall. 1956. Pp. xiv + 391. 52*s*.

Wahrscheinlichkeitstheorie. By H. RICHTER. Berlin: Springer-Verlag. 1956. Pp. xi + 435. DM.66.

Indice Cronologico de Legislacion Estadistica 1813–1956.
Resumen historico de la Estadistica en Espana.
Publicaciones estadisticas de Espana.
Published on the occasion of the Centenary of the Estadistica Oficial Espanola, 1856–1956, by Instituto Nacional de Estadistica, Madrid.

Vocabulario Brasileiro de Estatistica. By MILTON DA SILVA RODRIGUES. Brasil: Universidade de Sao Paulo. 1956. Pp. 304.

Petrographic Model Analysis. By F. CHAYES. New York: John Wiley and Sons Inc.; London: Chapman and Hall. 1956. Pp. xii + 113. 44*s*.

Serie Cicliche ed Oscillanti. By FRANCESCO BRAMBILLA. Milan: Centro per la Ricerca Operativa, Universita Commerciale 'Luigi Bocconi'. 1955. Pp. 320.

Statistica. By FRANCESCO BRAMBILLA. Italy: La Golliárdica. Vol. I, La Variabilità Strutturale. 1955. Pp. 672. Vol. II, La Teoria Della Stima. 1956. Pp. 688. Price together L. 9000.

Scientific Inference (second edition). By Sir HAROLD JEFFREYS. Cambridge University Press. 1957. Pp. 236. 25*s*.

Culture and the Structural Evolution of the Neural System. [James Arthur Lecture on the Evolution of the Human Brain 1955.] New York: The American Museum of Natural History. 1956. Pp. 57.

The Lognormal Distribution, with special reference to its uses in economics. [University of Cambridge Department of Applied Economics Monograph 5.] By J. AITCHISON and J. A. C. BROWN. Cambridge University Press. 1957. Pp. 176. 35*s*.

TRACTS FOR COMPUTERS

Department of Statistics, University College, London

I. **Tables of the Digamma and Trigamma Functions.** By ELEANOR PAIRMAN, M.A.
Tables for summing $S = \sum_{i=1}^{\infty} \frac{1}{(p_1 i+q_1)(p_2 i+q_2)\dots(p_n i+q_n)}$ where the p's and q's are numerical factors. *Price 5s. net.*

V. **Table of Coefficients of Everett's Central-Difference Interpolation Formula.** By A. J. THOMPSON, PH.D. *Second edition. Price 7s. 6d. net.*

VIII. **Table of the Logarithms of the Complete Γ-Function (to ten decimal places) for Argument 2 to 1200 beyond Legendre's Range (Argument 1 to 2).** By EGON S. PEARSON, D.SC. *Price 5s. net.*

IX. **Log Γ (x) from $x = 1$ to 50·9 by intervals of 0·01.** By JOHN BROWNLEE, M.D., D.SC. *Price 5s. net.*

X. **On Quadrature and Cubature or on Methods of Determining Approximately Single and Double Integrals.** By J. O. IRWIN, D.SC. *Price 7s. 6d. net.*

XII. **Tables of the Probable Error of the Coefficient of Correlation.** By KARL HOLZINGER, PH.D. *Price 5s. net.*

XIII. **Bibliotheca Tabularum Mathematicarum,** being a Descriptive Catalogue of Mathematical Tables. Part I. A, Logarithms of Numbers. By JAMES HENDERSON, PH.D. *Price 9s. net.*

XV. **Random Sampling Numbers.** By L. H. C. TIPPETT, M.SC., with a Foreword by KARL PEARSON. *Price 5s. net.*

XXIII. **Tables of $\tan^{-1}x$ and $\log(1+x^2)$.** To assist in the calculation of the ordinates of a Pearson Type IV curve. By L. J. COMRIE, PH.D. *Price 5s. net.*

XXIV. **Random Sampling Numbers (2nd Series).** By M. G. KENDALL and B. BABINGTON SMITH. *Price 5s. net.*

XXV. **Random Normal Deviates.** By HERMAN WOLD. *Price 5s. net.*

XXVI. **Correlated Random Normal Deviates.** By E. C. FIELLER, T. LEWIS and E. S. PEARSON. *Price 10s. 6d. net.*

Nos. II, III, IV, VI and VII are out of print

LOGARITHMETICA BRITANNICA

A standard Table of Logarithms to Twenty Decimal Places. By A. J. THOMPSON, Ph.D. (commenced in 1922 to commemorate the tercentenary of the publication of HENRY BRIGGS' s *Arithmetica Logarithmica*).

The nine separate sections of this Table have now been issued, and the complete work consisting of the logarithms of numbers 10,000–100,000, together with Dr Thompson's General Introduction (98 pp.) is now available in two bound volumes.

Price £8. 8s. 0d.

Issued by the CAMBRIDGE UNIVERSITY PRESS, Bentley House, LONDON, N.W.1
on behalf of the
DEPARTMENT OF STATISTICS, UNIVERSITY COLLEGE, LONDON
and obtainable from any bookseller

(*All rights reserved*)

BIOMETRIKA Vol. 44, Parts 1 and 2

CONTENTS

PAGE

JOHN WISHART, 1898–1956 1

P. ARMITAGE. Restricted sequential procedures 9

M. S. BARTLETT. On theoretical models for competitive and predatory biological systems 27

V. T. PATIL. The consistency and adequacy of the Poisson–Markoff model for density fluctuations . . 43

E. J. HANNAN. Testing for serial correlation in least squares regression 57

S. KULLBACK and H. M. ROSENBLATT. On the analysis of multiple regression in k categories . . . 67

R. L. BROWN. Bivariate structural relation 84

JOHN W. WILKINSON. An analysis of paired comparison designs with incomplete repetitions 97

C. L. MALLOWS. Non-null ranking models. I 114

J. AITCHISON and S. D. SILVEY. The generalization of probit analysis to the case of multiple responses . . 131

S. C. PEARCE. Experimenting with organisms as blocks 141

D. R. COX. The use of a concomitant variable in selecting an experimental design 160

M. G. BULMER. Approximate confidence limits for components of variance 159

D. E. BARTON and F. N. DAVID. Multiple runs 168

D. V. LINDLEY. Binomial sampling schemes and the concept of information 179

D. V. LINDLEY. A statistical paradox 187

H. W. HASKEY. Stochastic cross-infection between two otherwise isolated groups 193

J. K. MACKENZIE and M. J. THOMSON. Some statistics associated with the random disorientation of cubes . 205

J. H. DARWIN. The difference between consecutive members of a series of random variables arranged in order of size 211

B. I. HARLEY. Relation between the distributions of non-central t and of a transformed correlation coefficient 219

A. CLIFFORD COHEN, JR. On the solution of estimating equations for truncated and censored samples from normal populations 225

F. G. FOSTER and D. H. REES. Upper percentage points of the generalized Beta distribution. I . . . 237

MISCELLANEA

E. S. PAGE. On problems in which a change in a parameter occurs at an unknown point 248

D. J. BARTHOLOMEW. Testing for departure from the exponential distribution 253

B. I. HARLEY and E. S. PEARSON. The distribution of range in normal samples with $n = 200$ . . . 257

M. G. KENDALL. Studies in the history of probability and statistics. V. A note on playing cards . . 260

ALAN STUART. A singularity in the estimation of binomial variance 262

AUREL WINTNER. Student's distribution and Riemann's elliptic geometry 264

JOHN GURLAND. Some interrelations among compound and generalized distributions 265

M. S. BARTLETT. A note on tests of significance for linear functional relationships 268

M. G. KENDALL. The moments of the Leipnik distribution 270

M. H. QUENOUILLE. The effect of transformations of variables upon their correlation coefficients . . 272

B. I. HARLEY. Further properties of an angular transformation of the correlation coefficient . . . 273

FRANKLIN A. GRAYBILL and JOHN LEROY FOLKS. Heterogeneity of error variances in a randomized block design 275

G. P. SILLITTO. An extension property of a class of balanced incomplete block designs 278

N. L. JOHNSON. Sequentially determined confidence intervals 279

H. A. DAVID. Estimation of means of normal populations from observed minima 282

J. H. DARWIN. The power of the Poisson index of dispersion 286

H. O. LANCASTER. Some properties of the bivariate normal distribution considered in the form of a contingency table 289

REVIEWS

L. SCHMETTERER'S 'Einführung in die mathematische Statistik' 293

SIR RONALD A. FISHER'S 'Statistical Methods and Scientific Inference' 293

H. A. MEYER (ed.), 'Symposium on Monte Carlo Methods' 295

F. G. CORNELL'S 'The Essentials of Educational Statistics' 296

W. A. NEISWANGER'S 'Elementary Statistical Methods' 296

W. ALLEN WALLIS and HARRY V. ROBERTS'S 'Statistics: a New Approach' 297

R. B. GOLDSCHMIDT'S 'Theoretical Genetics' 298

M. G. KENDALL'S 'Rank Correlation Methods' 298

A. N. KOLMOGOROV'S 'Probability' 298

O. PERRON'S 'Irrationalzahlen' 299

A. ZYGMUND'S 'Trigonometrical Series' 299

Publications of the Mathematical Institute of the Hungarian Academy of Sciences, vol. I 299

K. L. NIELSEN'S 'Methods in Numerical Analysis' 299

NATIONAL BUREAU OF STANDARDS, Publications of the U.S. Department of Commerce, Applied Mathematics Series 37, 39, 46 and Ordnance Corps Pamphlet PB 111389 300

CORRIGENDA 301

OTHER BOOKS RECEIVED 302

Printed in Great Britain at the University Press, Cambridge (Brooke Crutchley, University Printer)

AGE
1
9
27
43
57
67
84
97
114
131
141
150
159
168
179
187
193
205
211
219
225
237
248
253
257
260
262
264
265
268
270
272
273
275
278
279
282
286
289
293
293
295
296
296
297
298
298
298
299
299
299
299
300
301
303